

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 21, 2004, 12:26:11 ; Search time 100.28 Seconds
(without alignments)
1558.126 Million cell updates/sec

Title: US-09-981-649A-24
Perfect score: 3060
Sequence: 1 MFLPWSLALPLLSSWVAGF.....VDGVLVSLGCPDLSLVDD 553

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3060	100.0	553	3 AAB01423	Rab01423 Human TAN
2	3060	100.0	553	5 AAO15368	Aao15368 Human EGF
3	3060	100.0	553	5 AAE26506	Aae26506 Human epi
4	3060	100.0	553	5 ABj05586	Abj05586 Breast ca
5	3060	100.0	553	6 ABg72942	Abg72942 Novel hum
6	3060	100.0	553	6 ABR48234	Abr48234 Human bla
7	3060	100.0	553	6 ABu62265	Abu62265 Lung canc
8	3060	100.0	553	6 ABu62265	Abu62265 Epidermal
9	3060	100.0	553	7 ADB80482	Adb80482 Ovarian c
10	3055	99.8	553	4 AAM93622	Aam93622 Human pol
11	3050	99.7	553	5 AAO15361	Aao15361 Human EGF
12	3050	99.7	553	5 AAE26500	Aae26500 Human epi
13	3050	99.7	553	6 ABG72935	Abg72935 Novel hum
14	3050	99.7	553	6 ABu62258	Abu62258 Epidermal
15	3049.5	99.7	554	4 AAB27224	Aab27224 Human EXM
16	3049.5	99.7	554	4 AAM39156	Aam39156 Human pol
17	3049.5	99.7	554	5 AAO15371	Aao15371 Human EGF
18	3049.5	99.7	554	5 AAO15370	Aao15370 Human EGF
19	3049.5	99.7	554	6 ABG72945	Abg72945 Novel hum
20	3049.5	99.7	554	6 ABG72944	Abg72944 Novel hum
21	3049.5	99.7	554	6 ABU62267	Abu62267 Novel epi
22	3049.5	99.7	554	6 ABU62268	Abu62268 Novel epi
23	3047	99.6	559	5 AAO15369	Aao15369 Human EGF
24	3047	99.6	559	6 ABG72943	Abg72943 Novel hum
25	3047	99.6	559	6 ABU62266	Abu62266 Novel epi

26	3030	99.0	553	2 AAY18108	Aay18108 Protein e
27	2986.5	97.6	573	4 AAM40942	Aam40942 Human pol
28	2769	90.5	537	5 AAO15360	Aao15360 Human EGF
29	2769	90.5	537	5 AAE26498	Aae26498 Human epi
30	2769	90.5	537	6 ABG72934	Abg72934 Novel hum
31	2769	90.5	537	6 ABU62257	Abu62257 Epidermal
32	2767	90.4	537	2 AAY18110	Aay18110 Protein e
33	2757	90.1	502	5 AAO15367	Aao15367 Human EGF
34	2757	90.1	502	5 AAE26499	Aae26499 Human epi
35	2757	90.1	502	6 ABG72941	Abg72941 Novel hum
36	2757	90.1	502	6 ABU62264	Abu62264 Epidermal
37	2438	79.3	551	5 ABB72294	Abb72294 Rat prote
38	1931	63.1	338	2 AAY41702	Aay41702 Human PRO
39	1931	63.1	338	3 AAB44258	Aab44258 Human PRO
40	1931	63.1	338	3 AAY95339	Aay95339 Human PRO
41	1931	63.1	338	3 AAB18669	Aab18669 Anino aci
42	1931	63.1	338	6 ABO25204	Abo25204 Novel hum
43	1931	63.1	338	6 ABU72210	Abu72210 Novel hum
44	1931	63.1	338	6 ABU84890	Abu84890 Human sec
45	1931	63.1	338	6 ABU61088	Abu61088 Human PRO

ALIGNMENTS

RESULT 1
AAB01423
ID AAB01423 standard; protein; 553 AA.
XX
XX AAB01423;
AC
XX
DT 20-OCT-2000 (first entry)
XX
DE Human TANGO 212.
XX
KW TANGO; 128; 140; 197; 212; 213; 239; modulating agent; asthma;
KW graft versus-host diseases; rheumatoid arthritis; psoriasis;
KW inflammatory bowel disease; septic shock; ulcerative colitis;
KW Crohn's disease; chronic myelogenous leukemia; cancer; liver disease;
KW Hodgkin's disease; osteoarthritis; Lyme's disease; cachexia;
KW autoimmune disease; myasthenia gravis; autoimmune diabetes;
KW systemic lupus erythematosus; transgenic animal; diagnosis; prognosis;
KW prophylactic; therapeutic; human.
XX
XX Homo sapiens.
XX
XX WO200039284-A1.
XX
XX 06-JUL-2000.
XX
XX 23-DEC-1999; 99WO-US031025.
XX
XX 30-DEC-1998; 98US-00223546.
XX
XX (MILL-) MILLENNIUM PHARM INC.
XX
XX Holtzman DA;
XX
XX WPI; 2000-465743/40.
XX
XX N-PSDB; AAA47456.
XX
XX Novel nucleic acid sequences encoding TANGO-128, 140, 197, 212, 213, 224
XX and 239 polypeptides useful for the treatment of asthma, rheumatoid
XX arthritis, psoriasis and autoimmune diseases.
XX
XX Claim 8; Fig 5; 209pp; English.
XX
XX Nucleic acids encoding TANGO polypeptides are useful as modulating agents
XX for regulating cellular processes like asthma, graft versus-host
XX diseases, rheumatoid arthritis, psoriasis, inflammatory bowel disease,
XX septic shock, ulcerative colitis, Crohn's disease, chronic myelogenous
XX leukemia, cancer, liver disease, Hodgkin's disease, osteoarthritis,
XX Lyme's disease, cachexia and autoimmune diseases e.g. myasthenia gravis,

CC autoimmune diabetes and systemic lupus erythematosus. The nucleic acids
CC are also useful for producing transgenic animals and the TANGO
CC polypeptides themselves. Partial TANGO-128, 140, 197, 212, 224, 239
CC sequences are useful in forensic biology, for diagnostic assays,
CC prognostic assays, pharmacogenomics and for monitoring clinical trials.
CC TANGO polypeptides are suitable for both prophylactic and therapeutic
CC methods for treating a subject at risk of a disorder or having a disorder
CC associated with aberrant TANGO expression. A wide range of cellular
CC disorders can be treated
XX
XX
XX Sequence 553 AA;

Query Match 100.0%; Score 3060; DB 3; Length 553;
Best Local Similarity 100.0%; Pred. No. 7.1e-213;
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MPLPWSLALPLLSSWVAGFGNAAARHGLLASARPGVCHYGTKLACCYGWRRNSKGV 60
Db 1 MPLPWSLALPLLSSWVAGFGNAAARHGLLASARPGVCHYGTKLACCYGWRRNSKGV 60

Qy 61 CEATCEPGCKGECVGNPKRCFPYGTGKTCSDQVNECGMKPRPCQHRVCVNTGSHYKFC 120
Db 61 CEATCEPGCKGECVGNPKRCFPYGTGKTCSDQVNECGMKPRPCQHRVCVNTGSHYKFC 120

Qy 121 LSGHMLMPDATCVNSRTCAMINQYSCDTEEGPQCLCPSSGLRLAPNGRDCLDIDECAS 180
Db 121 LSGHMLMPDATCVNSRTCAMINQYSCDTEEGPQCLCPSSGLRLAPNGRDCLDIDECAS 180

Qy 181 GKVICPNRRVCVNTFGSYCKCHIGPELQYISGRYDCIDINECTMDSHTCSHHANCNTQ 240
Db 181 GKVICPNRRVCVNTFGSYCKCHIGPELQYISGRYDCIDINECTMDSHTCSHHANCNTQ 240

Qy 241 GSFCKCKQYKGNGLRCSAIPENSVEVLRAFGTIKDRIKKLAHNSMCKKAKIKNVT 300
Db 241 GSFCKCKQYKGNGLRCSAIPENSVEVLRAFGTIKDRIKKLAHNSMCKKAKIKNVT 300

Qy 301 PEPTRTPTPKVNLQPFNYEEIVSRGNSHGKGNKEEKKEGLEDEKREKALKNDIEER 360
Db 301 PEPTRTPTPKVNLQPFNYEEIVSRGNSHGKGNKEEKKEGLEDEKREKALKNDIEER 360

Qy 361 SLRGDVFPPKVNAGEFGLILVQRKALTSLKLEHKLNLISVDCSFNHHGICDWKQREDDFD 420
Db 361 SLRGDVFPPKVNAGEFGLILVQRKALTSLKLEHKLNLISVDCSFNHHGICDWKQREDDFD 420

Qy 421 WNPADRDNAIGFYNAVPALAGHKDIDGLKLLPDLQPSNFCFLFDYRLAGDKVGLRV 480
Db 421 WNPADRDNAIGFYNAVPALAGHKDIDGLKLLPDLQPSNFCFLFDYRLAGDKVGLRV 480

Qy 481 FVXNSNNALAWKTTSEDEKWKTKGIQLYQGTDTATKSIIFAEARGKGTGIAVDGVLLV 540
Db 481 FVXNSNNALAWKTTSEDEKWKTKGIQLYQGTDTATKSIIFAEARGKGTGIAVDGVLLV 540

Qy 541 SGLCPDLSLLSVD 553
Db 541 SGLCPDLSLLSVD 553

RESULT 2
AAO15368
ID AAO15368 standard; protein; 553 AA.
XX AAO15368;
XX
XX
DT 19-SEP-2002 (first entry)
XX Human EGF motif-containing protein, SEQ ID No 24.
DE Human; epidermal growth factor motif; EGF motif; EGF6;
KW epithelial tissue growth; tissue repair; tissue regeneration;
KW corneal transplant healing; skin graft; wound healing; cancer; leukaemia;
KW nervous system disorder; infection; autoimmune disorder; inflammation;
KW multiple sclerosis; anaemia; periodontal disease; haemophilia;
KW fertility enhancement.

XX Homo sapiens.
XX WO2002030977-A2.
XX 18-APR-2002.
XX 15-OCT-2001; 2001WO-US032257.
XX 13-OCT-2000; 2000US-00687860.
XX (HYSE-) HYSEQ INC.
XX Asundi V, Ford JE, Drmanac RT, Liu C, Yamasaki V, Yeung G;
PI Tang TY, Zhang J, Zhou P, Zhou H;
XX WPI; 2002-426270/45.
XX N-PSDB; AAL43901.
XX Novel isolated epidermal growth factor motif polypeptide, termed EGF6,
PT for treating cancer, nervous system disorders, immune deficiencies,
PT autoimmune disorders, coagulation disorders and inflammatory conditions.
XX Claim 28; Page 167-169; 183pp; English.
XX The invention comprises the amino acid and coding sequences of human
CC epidermal growth factor (EGF) motif-containing proteins (EGF6 proteins).
CC The DNA and protein sequences of the invention are useful for inhibiting
CC the proliferation of cells expressing an EGF6 protein. The DNA and
CC protein sequences of the invention are useful for stimulating epithelial
CC tissue growth, for tissue repair and regeneration, corneal transplant
CC healing, skin graft production and wound healing. The DNA and protein
CC sequences are useful for treating cancer, leukaemia, nervous system
CC disorders, infection, autoimmune disorders (e.g. multiple sclerosis),
CC anaemia, periodontal diseases, haemophilia, inflammatory conditions, and
CC for effecting bodily characteristics and fertility of male or female
CC subjects. The present amino acid sequence represents a human EGF motif-
CC containing protein
XX
XX Sequence 553 AA;

Query Match 100.0%; Score 3060; DB 5; Length 553;
Best Local Similarity 100.0%; Pred. No. 7.1e-213;
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MPLPWSLALPLLSSWVAGFGNAAARHGLLASARPGVCHYGTKLACCYGWRRNSKGV 60
Db 1 MPLPWSLALPLLSSWVAGFGNAAARHGLLASARPGVCHYGTKLACCYGWRRNSKGV 60

Qy 61 CEATCEPGCKGECVGNPKRCFPYGTGKTCSDQVNECGMKPRPCQHRVCVNTGSHYKFC 120
Db 61 CEATCEPGCKGECVGNPKRCFPYGTGKTCSDQVNECGMKPRPCQHRVCVNTGSHYKFC 120

Qy 121 LSGHMLMPDATCVNSRTCAMINQYSCDTEEGPQCLCPSSGLRLAPNGRDCLDIDECAS 180
Db 121 LSGHMLMPDATCVNSRTCAMINQYSCDTEEGPQCLCPSSGLRLAPNGRDCLDIDECAS 180

Qy 181 GKVICPNRRVCVNTFGSYCKCHIGPELQYISGRYDCIDINECTMDSHTCSHHANCNTQ 240
Db 181 GKVICPNRRVCVNTFGSYCKCHIGPELQYISGRYDCIDINECTMDSHTCSHHANCNTQ 240

Qy 241 GSFCKCKQYKGNGLRCSAIPENSVEVLRAFGTIKDRIKKLAHNSMCKKAKIKNVT 300
Db 241 GSFCKCKQYKGNGLRCSAIPENSVEVLRAFGTIKDRIKKLAHNSMCKKAKIKNVT 300

Qy 301 PEPTRTPTPKVNLQPFNYEEIVSRGNSHGKGNKEEKKEGLEDEKREKALKNDIEER 360
Db 301 PEPTRTPTPKVNLQPFNYEEIVSRGNSHGKGNKEEKKEGLEDEKREKALKNDIEER 360

Qy 361 SLRGDVFPPKVNAGEFGLILVQRKALTSLKLEHKLNLISVDCSFNHHGICDWKQREDDFD 420
Db 361 SLRGDVFPPKVNAGEFGLILVQRKALTSLKLEHKLNLISVDCSFNHHGICDWKQREDDFD 420

QY 421 WNPADRDNAIGFYMAVPALAGHKDGRKLLKLLPOLQPOSNFCLLFDYRLAGDKVGLRV 480
DB 421 WNPADRDNAIGFYMAVPALAGHKDGRKLLKLLPOLQPOSNFCLLFDYRLAGDKVGLRV 480
QY 481 FVKNNSNALAWEXTTSEDEKWKTKIQLYQGTDTATKSIIFEAERGKGTGEIADVGLLV 540
DB 481 FVKNNSNALAWEXTTSEDEKWKTKIQLYQGTDTATKSIIFEAERGKGTGEIADVGLLV 540
QY 541 SGLCPDLSLVSDD 553
DB 541 SGLCPDLSLVSDD 553

RESULT 3
AAE26506
ID AAE26506 standard; protein; 553 AA.

XX AAE26506;
XX 13-DEC-2002 (first entry)
XX Human epidermal growth factor (EGF)-repeat containing protein #5.
XX Human; antibody; epidermal growth factor; EGF repeat; brain tumour;
KW nervous disorder; ulcer; leukaemia.
XX Homo sapiens.

Key Location/Qualifiers
FT Peptide 1..21
FT /label= Signal_peptide
FT Protein 22..553
FT /note= "Mature human EGF-repeat containing protein"
FT Domain 80..93
FT /note= "EGF motif 1"
FT Domain 95..128
FT /note= "EGF motif 2"
FT Domain 133..168
FT /note= "EGF motif 3"
FT Domain 175..214
FT /note= "EGF motif 4"
FT Domain 220..259
FT /note= "EGF motif 5"
FT Modified-site 247
FT /note= "N-glycosylation site"
FT Modified-site 346
FT /note= "N-glycosylation site"
FT Domain 363..365
FT /note= "RGD motif"
FT Domain 446..465
FT /note= "Transmembrane domain"
FT Modified-site 509
FT /note= "Tyrosine phosphorylation site"

US6392019-B1.
XX 21-MAY-2002.
XX 28-JUL-1999; 99US-00363316.
XX 22-NOV-1997; 97US-00968800.
PR 12-FEB-1999; 99US-00249697.
XX (FORD/) FORD J.
PA (YEUN/) YEUNG G.
XX Ford J, Yeung G;
XX WPI; 2002-424836/45.
DR N-PSDB; AAD44343.
XX Novel antibody specific for an epidermal growth factor repeat-containing
PT polypeptide, useful for the diagnosis of brain tumors, ulcers, leukemias,

PT and nervous disorders.
XX Example 4; Col 89-92; 92pp; English.
XX The invention relates to an antibody specific for a 537 residue epidermal
CC growth factor (EGF) repeat-containing polypeptide sequence. The invention
CC is used for detecting the presence of EGF repeat containing polypeptides
CC in a sample, in the diagnosis of brain tumours, nervous disorders,
CC ulcers, and leukaemias. The present sequence is human EGF-repeat
CC containing protein
XX Sequence 553 AA;
SQ

Query Match 100.0%; Score 3060; DB 5; Length 553;
Best Local Similarity 100.0%; Pred. No. 7.1e-213;
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MPLPWSLALPLLLSWVAGFGNAASARHGLLASARQPGVCHYGTKLACCYGWRNSKGV 60
DB 1 MPLPWSLALPLLLSWVAGFGNAASARHGLLASARQPGVCHYGTKLACCYGWRNSKGV 60
QY 61 CEATCEPGCKGECVGNKCRCPGVTGKTSQDYNCEGKPRPCQHRVCVNTHSGSKFC 120
DB 61 CEATCEPGCKGECVGNKCRCPGVTGKTSQDYNCEGKPRPCQHRVCVNTHSGSKFC 120
QY 121 LSGHMLMPDATCVNSRTCAMINCOYSCDETEGPOCLPSSGLRLAPNGRDCLDIDECAS 180
DB 121 LSGHMLMPDATCVNSRTCAMINCOYSCDETEGPOCLPSSGLRLAPNGRDCLDIDECAS 180
QY 181 GKVICPNRRVCNTFGSYCKCHIGFELYISGRYDCIDINECTWDSHTCSHHANCFTQ 240
DB 181 GKVICPNRRVCNTFGSYCKCHIGFELYISGRYDCIDINECTWDSHTCSHHANCFTQ 240
QY 241 GSFCKCKOGYKNGLRCSAIPENSVKELRAPGTIKDKRIKLLAHKNSMKKKAKIKNT 300
DB 241 GSFCKCKOGYKNGLRCSAIPENSVKELRAPGTIKDKRIKLLAHKNSMKKKAKIKNT 300
QY 301 PEPTRTPTPKVNLQPNFYEEIIVSRGNSHGKKGNEEKMEKLEDEKREKALKNDIEER 360
DB 301 PEPTRTPTPKVNLQPNFYEEIIVSRGNSHGKKGNEEKMEKLEDEKREKALKNDIEER 360
QY 361 SLRGDVFPPKVNAGEFGLILVORKALTSKLEHKDLNISVDCSFNHI COWKQDRDDFD 420
DB 361 SLRGDVFPPKVNAGEFGLILVORKALTSKLEHKDLNISVDCSFNHI COWKQDRDDFD 420
QY 421 WNPADRDNAIGFYMAVPALAGHKDGRKLLKLLPOLQPOSNFCLLFDYRLAGDKVGLRV 480
DB 421 WNPADRDNAIGFYMAVPALAGHKDGRKLLKLLPOLQPOSNFCLLFDYRLAGDKVGLRV 480
QY 481 FVKNNSNALAWEXTTSEDEKWKTKIQLYQGTDTATKSIIFEAERGKGTGEIADVGLLV 540
DB 481 FVKNNSNALAWEXTTSEDEKWKTKIQLYQGTDTATKSIIFEAERGKGTGEIADVGLLV 540
QY 541 SGLCPDLSLVSDD 553
DB 541 SGLCPDLSLVSDD 553

RESULT 4
ABJ05586
ID ABJ05586 standard; protein; 553 AA.
XX AC ABJ05586;
XX 14-NOV-2002 (first entry)
DT Breast cancer-associated protein 51.
XX Breast cancer; breast cancer-associated gene sequence; drug development;
XX pharmacogenetics; biosensor development.
XX Unidentified.
OS
XX

PN WO200259377-A2.
 XX
 PD 01-AUG-2002.
 XX
 PF 24-JAN-2002; 2003WO-US002242.
 XX
 PR 24-JAN-2001; 2001US-0263965P.
 PR 02-FEB-2001; 2001US-0265928P.
 PR 09-APR-2001; 2001US-00829472.
 PR 09-APR-2001; 2001US-0282698P.
 PR 04-MAY-2001; 2001US-0288590P.
 PR 29-MAY-2001; 2001US-0294443P.
 XX
 PA (EOSB-) EOS BIOTECHNOLOGY INC.
 XX
 PI Mack DH, Gish KC, Afar D;
 XX
 DR WPI; 2002-583738/62.
 DR N-PSDB; ABT07743.
 XX
 PT Detecting a breast cancer-associated transcript in a patient's cell,
 PT useful for diagnosing breast cancer, comprises contacting a biological
 PT sample with a polynucleotide that selectively hybridizes with breast
 PT cancer nucleic acids.
 XX
 PS Disclosure; Page 393; 414pp; English.
 XX
 CC The invention comprises a method of detecting a breast cancer-associated
 CC transcript in a cell from a patient. The method of the invention involves
 CC contacting a biological sample from the patient with a nucleotide that
 CC hybridizes to one of the 69 breast cancer-associated gene sequences shown
 CC in the specification. The method of the invention is useful in the
 CC diagnosis or prognosis of breast cancer, and for detecting genes that are
 CC up or down-regulated in breast cancer cells. Genes identified by the
 CC method of the invention can be used in diagnostic purposes and also as
 CC targets for screening for therapeutic compounds that modulate breast
 CC cancer (e.g. hormones or antibodies). Identification of genes that are
 CC over or under expressed in breast cancer can additionally provide high-
 CC resolution, high-sensitivity datasets which can be used in the areas of
 CC diagnostics, therapeutics, drug development, pharmacogenetics, protein
 CC structure and biosensor development. Amino acid sequences ABJ05536 -
 CC ABJ05604 represent the proteins encoded by the 69 breast cancer-
 CC associated genes of the invention
 XX
 SQ Sequence 553 AA;
 Query Match 100.0%; Score 3060; DB 5; Length 553;
 Best Local Similarity 100.0%; Pred. No. 7.1e-213; Indels 0; Gaps 0;
 Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MFLPWSLALPLLSSWAGGFGNNAARHGLLASARQPGVCHYGTKLACCYGMWRNSKGV 60
 DB 1 MFLPWSLALPLLSSWAGGFGNNAARHGLLASARQPGVCHYGTKLACCYGMWRNSKGV 60
 QY 61 CEATCEPGCKGECVGNPKRCFPGYGTGTCQDVNECGMKPRPCQHRVCVNTGSKYKFC 120
 DB 61 CEATCEPGCKGECVGNPKRCFPGYGTGTCQDVNECGMKPRPCQHRVCVNTGSKYKFC 120
 QY 121 LSGHMLMPDATCVNRTCAINQVSCDETEGPOCLPSSGLRLAPNGRCLDIDECAS 180
 DB 121 LSGHMLMPDATCVNRTCAINQVSCDETEGPOCLPSSGLRLAPNGRCLDIDECAS 180
 QY 181 GKVICPNRRCVNTFGSYKCHIGFELQYISGRYDCIDINECTWDSHTCSHHANCFNTQ 240
 DB 181 GKVICPNRRCVNTFGSYKCHIGFELQYISGRYDCIDINECTWDSHTCSHHANCFNTQ 240
 QY 241 GSFKCKCKQYKGNGLRCSAIPENSVKVLRAPGTIKDKIKLLAHKSMKKAKIKNVT 300
 DB 241 GSFKCKCKQYKGNGLRCSAIPENSVKVLRAPGTIKDKIKLLAHKSMKKAKIKNVT 300
 QY 301 PEPTRTPTPKVNLQPPNYEIVSRGNSHGKKGNEKMEKGELEKREKALKNDIBER 360
 DB 301 PEPTRTPTPKVNLQPPNYEIVSRGNSHGKKGNEKMEKGELEKREKALKNDIBER 360

QY 361 SLRGDVFPPKNEAGEFGLILVQRKALTSKLEHKDLNISVDCSFNHGICDWKQDREDDFD 420
 DB 361 SLRGDVFPPKNEAGEFGLILVQRKALTSKLEHKDLNISVDCSFNHGICDWKQDREDDFD 420
 QY 421 WNPADRDNAIGFYMAVPALAGHKDIGRLKLLLPDLOPQSNFCLLFDYRLAGDKVGLRV 480
 DB 421 WNPADRDNAIGFYMAVPALAGHKDIGRLKLLLPDLOPQSNFCLLFDYRLAGDKVGLRV 480
 QY 481 FVKNSSNALAWKTTSEDEKWKTKIQLYQGTDTATKSIIFEAERGKGTGSIADVGLLV 540
 DB 481 FVKNSSNALAWKTTSEDEKWKTKIQLYQGTDTATKSIIFEAERGKGTGSIADVGLLV 540
 QY 541 SGLCPDSLLSVDD 553
 DB 541 SGLCPDSLLSVDD 553
 RESULT 5
 ABG72942
 ID ABG72942 standard; protein; 553 AA.
 XX
 AC ABG72942;
 XX
 DT 02-APR-2003 (first entry)
 XX
 DE Novel human EGF-motif containing protein EGFL6.
 XX
 KW EGF; epidermal growth factor; cancer; lung cancer; brain cancer;
 KW prostate cancer; breast cancer; skin cancer; lymphoma cancer;
 KW sarcoma cancer; colon cancer; tumorigenicity; tumour site reduction;
 KW cell proliferation inhibition; vaccine; antisense gene therapy; human;
 KW EGFL6.
 XX
 OS Homo sapiens.
 XX
 FN US2002132250-A1.
 XX
 PD 19-SEP-2002.
 XX
 PF 15-OCT-2001; 2001US-00981649.
 XX
 PR 28-JUL-1999; 99US-00363316.
 PR 13-OCT-2000; 2000US-00687860.
 XX
 PA (FORD/) FORD J E.
 PA (YEUN/) YEUNG G.
 PA (ZHOU/) ZHOU H.
 XX
 PI Ford JE, Yeung G, Zhou H;
 XX
 DR WPI; 2003-174078/17.
 DR N-PSDB; ABX14779.
 XX
 PT Detecting cancerous cells expressing polynucleotides/polypeptides in
 PT samples, by contacting samples with labeled polynucleotides complementary
 PT to polynucleotide or an antibody against the polypeptide and detecting
 PT complex formed.
 XX
 PS Claim 13; Page 57-58; 78pp; English.
 XX
 CC The invention describes a method of detecting a cancerous cell expressing
 CC a polynucleotide (I) or a polypeptide (II) in a biological sample,
 CC involving contacting the sample with a labelled polynucleotide
 CC complementary to (I) or an antibody or its fragment that specifically
 CC binds to (II), for a period sufficient to form a complex and detecting
 CC the complex, so that if a complex is detected, the cell is detected. The
 CC method is useful for detecting cancerous cell in a biological sample such
 CC as tissue, cell, blood, serum, lymphatic fluid, urine, and cerebrospinal
 CC fluid. The cancerous cell is from lung, brain, prostate, breast, skin,
 CC lymphoma, sarcoma and colon. Preferably the cancer cell is A549 cell,
 CC -7 cell or SK-N-Mc cell. PC1 and PC2 are useful for inhibiting EGFL6
 CC activity, inhibiting tumorigenicity, reducing tumour sites and inhibiting

CC proliferation of a cancer cell. This is the amino acid sequence of the
 CC novel human EGF (epidermal growth factor) motif containing protein EGFL6
 XX Sequence 553 AA;
 SQ
 Query Match 100.0%; Score 3060; DB 6; Length 553;
 Best Local Similarity 100.0%; Pred. No. 7.1e-213;
 Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MFLPWSLALPLLLSWAGGFGNAASARHGLLASARQPGVCHYGTKLACCYGWRNSKGV 60
 DB 1 MFLPWSLALPLLLSWAGGFGNAASARHGLLASARQPGVCHYGTKLACCYGWRNSKGV 60
 QY 61 CEATCEPGCKFGCEVGNKRCFPFGYGTCTSDVNECGMKPRPCQHRVTHGSKYKFC 120
 DB 61 CEATCEPGCKFGCEVGNKRCFPFGYGTCTSDVNECGMKPRPCQHRVTHGSKYKFC 120
 QY 121 LSGHMLMPDATCVNSRTCAMINCOYSCDETEGPGQCLCPSSGRLAPNGRDCLDIDECAS 180
 DB 121 LSGHMLMPDATCVNSRTCAMINCOYSCDETEGPGQCLCPSSGRLAPNGRDCLDIDECAS 180
 QY 181 GKVICPNRRCVNTFGSYCKCHIGPELOVTSGRYDCIDINECTMSHTCSHHANCFNTQ 240
 DB 181 GKVICPNRRCVNTFGSYCKCHIGPELOVTSGRYDCIDINECTMSHTCSHHANCFNTQ 240
 QY 241 GSPKCKCKQGYKNGRLCSAIPENSVEVLAPGTIKDRIKLLAHKNSMKKAKIKNVT 300
 DB 241 GSPKCKCKQGYKNGRLCSAIPENSVEVLAPGTIKDRIKLLAHKNSMKKAKIKNVT 300
 QY 301 PEPTPTPTPKVNLQPNFYIEIVSRGNSHGKGNBEKMEGLEDEKREKALNDIEER 360
 DB 301 PEPTPTPTPKVNLQPNFYIEIVSRGNSHGKGNBEKMEGLEDEKREKALNDIEER 360
 QY 361 SLRGDVFPPKVNAGEFGLILVORKALTSKLEHKDLNISVDCSFNHCIDCWKQDREDDFD 420
 DB 361 SLRGDVFPPKVNAGEFGLILVORKALTSKLEHKDLNISVDCSFNHCIDCWKQDREDDFD 420
 QY 421 WNPADRDNAIGFYMAVPALAGHKDIGRLKLLPDLQPSNFCLLFDYRLAGDKVGLRV 480
 DB 421 WNPADRDNAIGFYMAVPALAGHKDIGRLKLLPDLQPSNFCLLFDYRLAGDKVGLRV 480
 QY 481 FVKNNSNALAWKTTSEDEKWKTKIQLYQGTDATKSIIFEABRGKGTGEIIVDGVLLV 540
 DB 481 FVKNNSNALAWKTTSEDEKWKTKIQLYQGTDATKSIIFEABRGKGTGEIIVDGVLLV 540
 QY 541 SGLCPDLSLLSVDD 553
 DB 541 SGLCPDLSLLSVDD 553
 RESULT 6
 ABR48234
 ID ABR48234 standard; protein; 553 AA.
 XX
 AC ABR48234;
 XX
 DT 12-JUN-2003 (first entry)
 XX
 DE Human bladder cancer associated protein sequence SEQ ID NO:189.
 XX
 KW Human; bladder cancer; cytostatic; gene therapy; vaccine.
 XX
 OS Homo sapiens.
 XX
 PN W02003003906-A2.
 XX
 PD 16-JAN-2003.
 XX
 PF 03-JUL-2002; 2002WO-US021338.
 XX
 PR 03-JUL-2001; 2001US-0302814P.
 PR 03-AUG-2001; 2001US-0310099P.
 PR 08-NOV-2001; 2001US-0343705P.

PR 13-NOV-2001; 2001US-0350666P.
 PR 12-APR-2002; 2002US-0372246P.
 XX (EOSB-) EOS BIOTECHNOLOGY INC.
 PA Mack DH, Aziz N;
 XX WIPI; 2003-201532/19.
 DR N-PSDB; ACC51050.
 DR
 XX Detecting a bladder cancer-associated transcript in a cell from a
 PT patient, comprises contacting a biological sample from the patient with a
 PT bladder cancer-associated polynucleotide or antibody.
 XX
 PS Claim 10; Page 294; 307pp; English.
 XX
 CC The present invention describes a method for detecting a bladder cancer-
 CC associated transcript in a cell from a patient. The method comprises
 CC contacting a biological sample from the patient with a polynucleotide
 CC that selectively hybridizes to a sequence that is 80 % identical to a
 CC table of sequences (see ACC50951 to ACC51059). ACC50951 to ACC51059
 CC encode the human bladder cancer-associated proteins given in ABR48146 to
 CC ABR48242). Bladder cancer-associated sequences from the present invention
 CC have cytostatic activities, and can be used in antisense gene therapy and
 CC in vaccine production. The method can be used for detecting a bladder
 CC cancer-associated transcript in a cell from a patient. The method is
 CC useful in diagnosing or treating bladder cancer and in screening for
 CC compounds that modulate bladder cancer, such as hormones or antibodies.
 CC The nucleic acid molecules from the present invention may be used in
 CC various screening and diagnostic methods, and for gene therapy, vaccine
 CC and/or antisense/inhibition applications
 XX
 SQ Sequence 553 AA;
 Query Match 100.0%; Score 3060; DB 6; Length 553;
 Best Local Similarity 100.0%; Pred. No. 7.1e-213;
 Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MFLPWSLALPLLLSWAGGFGNAASARHGLLASARQPGVCHYGTKLACCYGWRNSKGV 60
 DB 1 MFLPWSLALPLLLSWAGGFGNAASARHGLLASARQPGVCHYGTKLACCYGWRNSKGV 60
 QY 61 CEATCEPGCKFGCEVGNKRCFPFGYGTCTSDVNECGMKPRPCQHRVTHGSKYKFC 120
 DB 61 CEATCEPGCKFGCEVGNKRCFPFGYGTCTSDVNECGMKPRPCQHRVTHGSKYKFC 120
 QY 121 LSGHMLMPDATCVNSRTCAMINCOYSCDETEGPGQCLCPSSGRLAPNGRDCLDIDECAS 180
 DB 121 LSGHMLMPDATCVNSRTCAMINCOYSCDETEGPGQCLCPSSGRLAPNGRDCLDIDECAS 180
 QY 181 GKVICPNRRCVNTFGSYCKCHIGPELOVTSGRYDCIDINECTMSHTCSHHANCFNTQ 240
 DB 181 GKVICPNRRCVNTFGSYCKCHIGPELOVTSGRYDCIDINECTMSHTCSHHANCFNTQ 240
 QY 241 GSPKCKCKQGYKNGRLCSAIPENSVEVLAPGTIKDRIKLLAHKNSMKKAKIKNVT 300
 DB 241 GSPKCKCKQGYKNGRLCSAIPENSVEVLAPGTIKDRIKLLAHKNSMKKAKIKNVT 300
 QY 301 PEPTPTPTPKVNLQPNFYIEIVSRGNSHGKGNBEKMEGLEDEKREKALNDIEER 360
 DB 301 PEPTPTPTPKVNLQPNFYIEIVSRGNSHGKGNBEKMEGLEDEKREKALNDIEER 360
 QY 361 SLRGDVFPPKVNAGEFGLILVORKALTSKLEHKDLNISVDCSFNHCIDCWKQDREDDFD 420
 DB 361 SLRGDVFPPKVNAGEFGLILVORKALTSKLEHKDLNISVDCSFNHCIDCWKQDREDDFD 420
 QY 421 WNPADRDNAIGFYMAVPALAGHKDIGRLKLLPDLQPSNFCLLFDYRLAGDKVGLRV 480
 DB 421 WNPADRDNAIGFYMAVPALAGHKDIGRLKLLPDLQPSNFCLLFDYRLAGDKVGLRV 480
 QY 481 FVKNNSNALAWKTTSEDEKWKTKIQLYQGTDATKSIIFEABRGKGTGEIIVDGVLLV 540
 DB 481 FVKNNSNALAWKTTSEDEKWKTKIQLYQGTDATKSIIFEABRGKGTGEIIVDGVLLV 540

QY 541 SGLCPDLSLLSVD 553
 DB 541 SGLCPDLSLLSVD 553

RESULT 7
 ABUS6725
 ID ABUS6725 standard; protein; 553 AA.
 XX
 AC ABUS6725;
 XX
 DT 02-APR-2003 (first entry)
 XX
 DE Lung cancer-associated polypeptide #318.
 XX
 KW Lung cancer-associated polypeptide; cytostatic; emphysema;
 KW antiinflammatory; antiasthmatic; non-small cell lung cancer; atelectasis;
 KW small cell lung cancer; benign lesion; precancerous lesion; bronchitis;
 KW chronic obstructive pulmonary disease; hypersensitivity pneumonitis;
 KW interstitial pulmonary fibrosis; fibrosis; asthma; bronchiectasis.
 XX
 OS Unidentified.
 XX
 FN WO200286443-A2.
 XX
 PD 31-OCT-2002.
 XX
 PF 18-APR-2002; 2002WO-US012476.
 XX
 PR 18-APR-2001; 2001US-0284770P.
 PR 10-MAY-2001; 2001US-0290492P.
 PR 09-NOV-2001; 2001US-0339245P.
 PR 13-NOV-2001; 2001US-0350666P.
 PR 29-NOV-2001; 2001US-0334370P.
 PR 12-APR-2002; 2002US-0372246P.
 XX
 PA (EOSB-) EOS BIOTECHNOLOGY INC.
 XX
 PI Aziz N, Murray R;
 XX
 DR WPI; 2003-093161/08.
 DR N-PSDB; ABX76454.
 XX
 XX Detecting a lung cancer-associated transcript in a cell from a patient
 PT for treating lung cancer, by contacting a biological sample from the
 PT patient with a polynucleotide that exhibits increased or decreased
 PT expression in lung cancer.
 XX
 PS Claim 27; Page 435; 453pp; English.

The invention relates to a method for detecting a lung cancer-associated transcript in a cell from a patient, comprising contacting a biological sample from the patient with a polynucleotide that selectively hybridises to a sequence that is at least 80 % identical to a gene that exhibits increased or decreased expression in lung cancer samples. Lung cancer-associated polynucleotides and polypeptides are used for identifying a compound that modulates a lung cancer-associated polypeptide, for inhibiting proliferation of a lung cancer-associated cell to treat lung cancer in a patient and for treating a mammal having lung cancer by administering a modulatory compound identified. The methods are useful for treating lung cancer, such as small cell lung cancer, non-small cell lung cancer or other benign or precancerous lesions, e.g. atelectasis, emphysema, bronchitis, chronic obstructive pulmonary disease, fibrosis, hypersensitivity pneumonitis, interstitial pulmonary fibrosis, asthma and bronchiectasis. The genes, polynucleotides and polypeptides are useful for diagnostic purposes and as targets for screening for therapeutic compounds that modulate lung cancer, such as antibodies. Sequences ABUS6408-ABUS6745 represent lung cancer-associated polypeptides of the invention

Sequence 553 AA;

Query Match 100.0%; Score 3060; DB 6; Length 553;
 Best Local Similarity 100.0%; Pred. No. 7.1e-213;
 Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPLPWSLALPILLSWVAGFGNAAARHGLLASARQPGVCHYGTKLACCGWRNSKGV 60
 DB 1 MPLPWSLALPILLSWVAGFGNAAARHGLLASARQPGVCHYGTKLACCGWRNSKGV 60

QY 61 CEATCEPGCKFGECVGNPKRCFPYGTGKTCSQDVNECGMKRPRCQHRCVNTHGSKYKFC 120
 DB 61 CEATCEPGCKFGECVGNPKRCFPYGTGKTCSQDVNECGMKRPRCQHRCVNTHGSKYKFC 120

QY 121 LSGHMLMPDATCVNSRTCAMINCOYSCDETEEGQCLCPSSGLRLAPNGRCLDIDECAS 180
 DB 121 LSGHMLMPDATCVNSRTCAMINCOYSCDETEEGQCLCPSSGLRLAPNGRCLDIDECAS 180

QY 181 GKVICPNRRCVNTFGSYCKCHIGFELQYISGRYDCIDINECTWDSHTCSHHANCFNTQ 240
 DB 181 GKVICPNRRCVNTFGSYCKCHIGFELQYISGRYDCIDINECTWDSHTCSHHANCFNTQ 240

QY 241 GSFCKCKQGYKGNGLRCSAIPENSVEVLRAPGTIKDRIKKLLAHKNSMKKAKIKNVT 300
 DB 241 GSFCKCKQGYKGNGLRCSAIPENSVEVLRAPGTIKDRIKKLLAHKNSMKKAKIKNVT 300

QY 301 PEPTRTPTPKVNLQPFNYEEIVSRGNSHGKKGNEEKKEGLEDEKREKALKNDIEER 360
 DB 301 PEPTRTPTPKVNLQPFNYEEIVSRGNSHGKKGNEEKKEGLEDEKREKALKNDIEER 360

QY 361 SLRGDVFPPKVNAGEFGLILVQRKALTSKLEHKDLNISVDCSFNHHGICDMKQDREDDFD 420
 DB 361 SLRGDVFPPKVNAGEFGLILVQRKALTSKLEHKDLNISVDCSFNHHGICDMKQDREDDFD 420

QY 421 WNPADRDNAIGFYMAVPALAGHKDIGHLLKLLPDLQPSNFCLLFDYRLAGDKVGLRV 480
 DB 421 WNPADRDNAIGFYMAVPALAGHKDIGHLLKLLPDLQPSNFCLLFDYRLAGDKVGLRV 480

QY 481 FVKSNNALAWKTTSEDEKWKTKIQLYQCTDTSIIPAEARGKGTGIAVDGVLV 540
 DB 481 FVKSNNALAWKTTSEDEKWKTKIQLYQCTDTSIIPAEARGKGTGIAVDGVLV 540

QY 541 SGLCPDLSLLSVD 553
 DB 541 SGLCPDLSLLSVD 553

RESULT 8
 ABUS62265
 ID ABUS62265 standard; protein; 553 AA.
 XX
 AC ABUS62265;
 XX
 DT 01-SEP-2003 (first entry)
 XX
 DE Epidermal growth factor motif protein EGFL6 #2.
 KW Human; epidermal growth factor motif protein; EGFL6; cytostatic;
 KW neuroprotective; antibacterial; antiparasitic; antilepemic;
 KW antifertility; EGF-Agonist; EGF-Antagonist; cell growth; cancer;
 KW neurodegenerative disorder; leukaemia; brain tumour; lung tumour;
 KW breast tumour; gastrointestinal tumour; skin tumour; prostate tumour;
 KW carcinoma; parasite; biorhythm; fertility; metabolism; catabolism;
 KW anabolism.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 39 /label= OTHER
 FT /note= "OTHER= any amino acid"
 FT Misc-difference 40 /label= OTHER
 FT /note= "OTHER= any amino acid"
 FT Misc-difference 41 /label= OTHER
 FT /note= "OTHER= any amino acid"

FT /label= OTHER
 FT /note= "OTHER= any amino acid"
 FT Misc-difference 45
 FT /label= OTHER
 FT /note= "OTHER= any amino acid"
 FT Misc-difference 46
 FT /label= OTHER
 FT /note= "OTHER= any amino acid"
 XX
 PN US2003036508-A1.
 XX
 XX 20-FEB-2003.
 XX
 XX 17-APR-2002; 2002US-00124986.
 XX
 XX 22-NOV-1997; 97US-00968800.
 PR 12-FEB-1999; 99US-00249697.
 PR 28-JUL-1999; 99US-00363316.
 PR 13-OCT-2000; 2000US-00687860.
 PR 15-OCT-2001; 2001US-00981649.
 XX
 PA (FORD/) FORD J.
 PA (YEUN/) YEUNG G.
 PA (ZHOU/) ZHOU H.
 XX
 XX Ford J, Yeung G, Zhou H;
 XX
 XX WPI; 2003-492123/46.
 DR N-PSDB; ACD25942.
 XX
 XX Stimulating cell growth by contacting the cell with an EGFL6 polypeptide,
 PT useful for the diagnosis and treatment of cancers and neurodegenerative
 PT disorders.
 PT
 XX Claim 6; Page 66-67; 86pp; English.
 PS
 XX The invention describes a method of stimulating cell growth comprising
 XX contacting the cell with an EGFL6 polypeptide having at least 90 %
 CC sequence identity to a 553 amino acid sequence (S1), given in the
 CC specification, or its variant and/or fragment lacking a C-terminal
 CC portion of the EGFL6 polypeptide. The methods and compositions of the
 CC present invention are useful for the diagnosis and treatment of cancers
 CC and neurodegenerative disorders by stimulating cell growth. The cancers
 CC include leukaemia, brain, lung, breast, gastrointestinal, skin and
 CC prostate tumours and carcinomas. They can also be used in inhibiting the
 CC growth of infectious agents and parasites, effecting bodily
 CC characteristics and biorhythms, effecting fertility, metabolism
 CC catabolism and anabolism of fats, vitamins, proteins and minerals, and
 CC effecting behavioural characteristics. This is the amino acid sequence of
 CC novel human epidermal growth factor motif protein EGFL6
 XX
 XX Sequence 553 AA;
 SQ
 Query Match 100.0%; Score 3060; DB 6; Length 553;
 Best Local Similarity 100.0%; Pred. No. 7.1e-213;
 Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 YQ 1 MPLPWSLALPLLWSVAGGFGNAAARHHGLASARQPGVCHYGTKLACCGWERNKGV 60
 DB 1 MPLPWSLALPLLWSVAGGFGNAAARHHGLASARQPGVCHYGTKLACCGWERNKGV 60
 YQ 61 CEATCEPGCKFGECVGNKRCFFPGYGTGKTSQDVNECGMKPRPCQHRVCNVNTHGSYKFC 120
 DB 61 CEATCEPGCKFGECVGNKRCFFPGYGTGKTSQDVNECGMKPRPCQHRVCNVNTHGSYKFC 120
 YQ 121 LSGHMLPDPATCVNSRTCAMINQCYSCEDTBEGPOCLCPSSGLRLAPNGRCLDIDECAS 180
 DB 121 LSGHMLPDPATCVNSRTCAMINQCYSCEDTBEGPOCLCPSSGLRLAPNGRCLDIDECAS 180
 YQ 181 GKVICPNRRCVNTFGSYCKHIGFLOYISGRYDCLDINECTMDSTCHSHANCNTQ 240
 DB 181 GKVICPNRRCVNTFGSYCKHIGFLOYISGRYDCLDINECTMDSTCHSHANCNTQ 240

QY 241 GSPFKCKQGYKGNGLRCSAIPENSVKVLRAPGTIKDKRIKLLAHKNSMKKAKIKNVT 300
 DB 241 GSPFKCKQGYKGNGLRCSAIPENSVKVLRAPGTIKDKRIKLLAHKNSMKKAKIKNVT 300
 QY 301 PEPTRTPTKVNLOPNFYBEIVSRGNSHGKKGNEEKMEGLEDEKREKALKNDIEER 360
 DB 301 PEPTRTPTKVNLOPNFYBEIVSRGNSHGKKGNEEKMEGLEDEKREKALKNDIEER 360
 QY 361 SLRGDVFPPKNEAGFGLILVORKALTSKLEHKDLNISVDCSFNHCIDWKQDREDDFD 420
 DB 361 SLRGDVFPPKNEAGFGLILVORKALTSKLEHKDLNISVDCSFNHCIDWKQDREDDFD 420
 QY 421 WNPADRDNAIGFYMAVPALAGHKKDI GRKLLLPDLQPOSNFCLLDYRLAGDKVKGLRV 480
 DB 421 WNPADRDNAIGFYMAVPALAGHKKDI GRKLLLPDLQPOSNFCLLDYRLAGDKVKGLRV 480
 QY 481 FVKNSNNALAWKTTSEDEKWKTKIOLYQGTDTATKSIIFEAERGKGTGEIAVDGVLV 540
 DB 481 FVKNSNNALAWKTTSEDEKWKTKIOLYQGTDTATKSIIFEAERGKGTGEIAVDGVLV 540
 QY 541 SGLCPDLSLSVDD 553
 DB 541 SGLCPDLSLSVDD 553
 RESULT 9
 ADB80482
 ID ADB80482 standard; protein; 553 AA.
 XX
 AC ADB80482;
 XX
 DT 04-DEC-2003 (first entry)
 XX
 DE Ovarian cancer-associated protein #23.
 XX
 KW cytostatic; gene therapy; vaccine; ovarian cancer; diagnosis;
 KW post-operative chemotherapy; radiation therapy; tumour prognosis;
 KW pre-cancerous lesion detection.
 XX
 OS Homo sapiens.
 XX
 PN W02002102235-A2.
 XX
 PD 27-DEC-2002.
 XX
 PF 18-JUN-2002; 2002WO-US019297.
 XX
 PR 18-JUN-2001; 2001US-0299234P.
 PR 27-AUG-2001; 2001US-0315287P.
 PR 05-SEP-2001; 2001US-0317544P.
 PR 13-NOV-2001; 2001US-0350666P.
 PR 12-APR-2002; 2002US-0372246P.
 XX
 XX (EOSB-) EOS BIOTECHNOLOGY INC.
 PA
 XX Mack DH, Gish KC;
 PI
 XX WPI; 2003-167431/16.
 DR N-PSDB; ADB80481.
 DR
 XX
 PT Detecting an ovarian cancer-associated transcript in a cell from a
 PT patient, comprises contacting a biological sample from the patient with a
 PT polynucleotide that hybridizes to an ovarian cancer gene.
 XX
 PS Claim 13; Page 290-291; 332pp; English.
 XX
 CC The invention relates to a method of detecting an ovarian cancer-
 CC associated transcript in a cell from a patient, by contacting a
 CC biological sample from the patient with a polynucleotide that selectively
 CC hybridizes to a sequence at least 80% identical to any one of 80
 CC nucleic acid sequences given in the specification. The method is useful
 CC in diagnosing ovarian cancer and in identifying and using agents and/or
 CC targets that inhibit ovarian cancer. The nucleic acid molecule,
 CC

CC polypeptide and the antibody may also be used in detecting ovarian
 CC cancers, monitoring and early detection of relapse following treatment,
 CC monitoring response to therapy, selecting patients for post-operative
 CC chemotherapy or radiation therapy, in selecting mode of therapy,
 CC determining tumour prognosis, early detection of pre-cancerous lesions,
 CC and as vaccines. This sequence corresponds to one of the proteins used
 CC for the detection method of the invention.

XX SQ Sequence 553 AA;

Query Match 100.0%; Score 3060; DB 7; Length 553;
 Best Local Similarity 100.0%; Pred. No. 7, 1e-213;
 Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPLPWSLALPLLSSWAGGFGNAASARHHGLLASARQPGVCHYGTKLACCYGMWRNSKGV 60
 DB 1 MPLPWSLALPLLSSWAGGFGNAASARHHGLLASARQPGVCHYGTKLACCYGMWRNSKGV 60

QY 61 CEATCEPGCKFGECVGNKRCPCFPYGTGKTCSDQVNECGMKPRPCQHRVCNTHSGYKFC 120
 DB 61 CEATCEPGCKFGECVGNKRCPCFPYGTGKTCSDQVNECGMKPRPCQHRVCNTHSGYKFC 120

QY 121 LSGHMLMPDATCVNSRTCAMINCOYSCDETEEGPQCLCPSSGLRLAPNGRDCLDIDECAS 180
 DB 121 LSGHMLMPDATCVNSRTCAMINCOYSCDETEEGPQCLCPSSGLRLAPNGRDCLDIDECAS 180

QY 181 GKVICPNRRVCNTPGSIYCKCHIGFELQYISGRYDCIDINECTMDSHTCSHHANCFTQ 240
 DB 181 GKVICPNRRVCNTPGSIYCKCHIGFELQYISGRYDCIDINECTMDSHTCSHHANCFTQ 240

QY 241 GSFCKCKQGYKGNGLRCSAIPENSVEKVLRAPTIKDRIKLLAHKNSMKKAKIKNVT 300
 DB 241 GSFCKCKQGYKGNGLRCSAIPENSVEKVLRAPTIKDRIKLLAHKNSMKKAKIKNVT 300

QY 301 PEPTRTPTKVNLPFPNYEIVSRGNSHGKKGNEEKMEKEGLEDEKREKALKNDIEER 360
 DB 301 PEPTRTPTKVNLPFPNYEIVSRGNSHGKKGNEEKMEKEGLEDEKREKALKNDIEER 360

QY 361 SLRGDVFFPKVNEAGEFGLILVQRKALTSKLEHKDLNISVDCSFNHGICDWKQDREDDFD 420
 DB 361 SLRGDVFFPKVNEAGEFGLILVQRKALTSKLEHKDLNISVDCSFNHGICDWKQDREDDFD 420

QY 421 WNPADRDNAIGFYMAVPALAGHKDDIGRLKLLPDLQPOSNFCLLPDYRLAGDKVGLRV 480
 DB 421 WNPADRDNAIGFYMAVPALAGHKDDIGRLKLLPDLQPOSNFCLLPDYRLAGDKVGLRV 480

QY 481 FVKNSNNALAWKTTSEDEKWKTKIQLYQGTDATKSIIFEARGKGTGEIADVGVLLV 540
 DB 481 FVKNSNNALAWKTTSEDEKWKTKIQLYQGTDATKSIIFEARGKGTGEIADVGVLLV 540

RESULT 10

AA093622
 ID AA093622 standard; protein; 553 AA.

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

PF 07-JUL-2000; 2000EP-00114089.
 XX 08-JUL-1999; 99JP-00194486.
 PR 11-JAN-2000; 2000JP-00118774.
 PR 02-MAY-2000; 2000JP-00183765.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otauki T, Koga H;
 XX
 DR WPI: 2001-524255/58.
 DR N-PSDB; AAK94555.
 XX
 PT 830 Primers useful for synthesizing full length cDNA clones and their use
 PT in genetic manipulation.
 XX
 PS Claim 8; SEQ ID NO 3456; 1380pp + Sequence Listing; English.
 XX
 CC The invention relates to primers for synthesising full length cDNA
 CC clones. 830 cDNA molecules encoding a human protein have been isolated
 CC and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have
 CC been determined. Primers for synthesising the full length cDNA are useful
 CC for clarifying the function of the protein encoded by the cDNA. The full
 CC length clones were obtained by construction of full length enriched cDNA
 CC libraries that were synthesised by the oligo-capping method. The primers
 CC enable the production of the full length cDNA easily without any special
 CC methods. The present sequence is a polypeptide encoded by a full length
 CC human cDNA of the invention. Note: The sequence data for this patent did
 CC not form part of the printed specification, but was obtained in CD-ROM
 CC format directly from EPO
 XX
 SQ Sequence 553 AA;

Query Match 99.8%; Score 3055; DB 4; Length 553;
 Best Local Similarity 99.8%; Pred. No. 1, 6e-212;
 Matches 552; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPLPWSLALPLLSSWAGGFGNAASARHHGLLASARQPGVCHYGTKLACCYGMWRNSKGV 60
 DB 1 MPLPWSLALPLLSSWAGGFGNAASARHHGLLASARQPGVCHYGTKLACCYGMWRNSKGV 60

QY 61 CEATCEPGCKFGECVGNKRCPCFPYGTGKTCSDQVNECGMKPRPCQHRVCNTHSGYKFC 120
 DB 61 CEATCEPGCKFGECVGNKRCPCFPYGTGKTCSDQVNECGMKPRPCQHRVCNTHSGYKFC 120

QY 121 LSGHMLMPDATCVNSRTCAMINCOYSCDETEEGPQCLCPSSGLRLAPNGRDCLDIDECAS 180
 DB 121 LSGHMLMPDATCVNSRTCAMINCOYSCDETEEGPQCLCPSSGLRLAPNGRDCLDIDECAS 180

QY 181 GKVICPNRRVCNTPGSIYCKCHIGFELQYISGRYDCIDINECTMDSHTCSHHANCFTQ 240
 DB 181 GKVICPNRRVCNTPGSIYCKCHIGFELQYISGRYDCIDINECTMDSHTCSHHANCFTQ 240

QY 241 GSFCKCKQGYKGNGLRCSAIPENSVEKVLRAPTIKDRIKLLAHKNSMKKAKIKNVT 300
 DB 241 GSFCKCKQGYKGNGLRCSAIPENSVEKVLRAPTIKDRIKLLAHKNSMKKAKIKNVT 300

QY 301 PEPTRTPTKVNLPFPNYEIVSRGNSHGKKGNEEKMEKEGLEDEKREKALKNDIEER 360
 DB 301 PEPTRTPTKVNLPFPNYEIVSRGNSHGKKGNEEKMEKEGLEDEKREKALKNDIEER 360

QY 361 SLRGDVFFPKVNEAGEFGLILVQRKALTSKLEHKDLNISVDCSFNHGICDWKQDREDDFD 420
 DB 361 SLRGDVFFPKVNEAGEFGLILVQRKALTSKLEHKDLNISVDCSFNHGICDWKQDREDDFD 420

QY 421 WNPADRDNAIGFYMAVPALAGHKDDIGRLKLLPDLQPOSNFCLLPDYRLAGDKVGLRV 480
 DB 421 WNPADRDNAIGFYMAVPALAGHKDDIGRLKLLPDLQPOSNFCLLPDYRLAGDKVGLRV 480

QY 481 FVKNSNNALAWKTTSEDEKWKTKIQLYQGTDATKSIIFEARGKGTGEIADVGVLLV 540
 DB 481 FVKNSNNALAWKTTSEDEKWKTKIQLYQGTDATKSIIFEARGKGTGEIADVGVLLV 540

QY 541 SGLCPDLSLSVDD 553
 Db 541 SGLCPDLSLSVDD 553

RESULT 11
 ID AA015361 standard; protein; 553 AA.
 XX AA015361;
 XX 19-SEP-2002 (first entry)
 XX Human EGF motif-containing protein, SEQ ID No 6.
 XX Human; epidermal growth factor motif; EGF motif; EGF6;
 KW epithelial tissue growth; tissue repair; tissue regeneration;
 KW corneal transplant healing; skin graft; wound healing; cancer; leukaemia;
 KW nervous system disorder; infection; autoimmune disorder; inflammation;
 KW multiple sclerosis; anaemia; periodontal disease; haemophilia;
 KW fertility enhancement.
 XX Homo sapiens.
 OS
 XX
 XX Key Location/Qualifiers
 FT Misc-difference 357 /note= "Encoded by WTA"
 FT
 XX WO200230977-A2.
 XX
 XX 18-APR-2002.
 XX
 XX 15-OCT-2001; 2001WO-US032257.
 PF
 XX 13-OCT-2000; 2000US-00687860.
 PR
 XX (HYSE-) HYSEQ INC.
 PA
 PI Asundi V, Ford JE, Drmanac RT, Liu C, Yamasaki V, Yeung G;
 PI Tang TY, Zhang J, Zhou P, Zhou H;
 XX WPI: 2002-426270/45.
 DR N-PSDB; AAL43890.
 XX
 PT Novel isolated epidermal growth factor motif polypeptide, termed EGF6,
 PT for treating cancer, nervous system disorders, immune deficiencies,
 PT autoimmune disorders, coagulation disorders and inflammatory conditions.
 XX
 PS Example 3; Fig 5; 183pp; English.
 XX
 CC The invention comprises the amino acid and coding sequences of human
 CC epidermal growth factor (EGF) motif-containing proteins (EGF6 proteins).
 CC The DNA and protein sequences of the invention are useful for inhibiting
 CC the proliferation of cells expressing an EGF6 protein. The DNA and
 CC protein sequences of the invention are useful for stimulating epithelial
 CC tissue growth, for tissue repair and regeneration, corneal transplant
 CC healing, skin graft production and wound healing. The DNA and protein
 CC sequences are useful for treating cancer, leukaemia, nervous system
 CC disorders, infection, autoimmune disorders (e.g. multiple sclerosis),
 CC anaemia, periodontal diseases, haemophilia, inflammatory conditions, and
 CC for effecting bodily characteristics and fertility of male or female
 CC subjects. The present amino acid sequence represents a human EGF motif-
 CC containing protein
 XX
 SQ Sequence 553 AA;

Query Match 99.7%; Score 3050; DB 5; Length 553;
 Best Local Similarity 99.6%; Pred. No. 3.8e-212;
 Matches 551; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MFLPWSLALPELLPWVAGFGNAAARHGLLASARQGVCHYGTKLACCYGWRRNSKV 60
 |||||
 61 CEATCEPGCKFGECVGNKCRCPFGYTGKTCSDVNECGMKRPFQCHRCVNTHGSYKFC 120
 |||||
 61 CEATCEPGCKFGECVGNKCRCPFGYTGKTCSDVNECGMKRPFQCHRCVNTHGSYKFC 120
 |||||
 121 LSGHMLPDCATVNSRTCAMINCOYSCDTEEGPQCLCPSSGLRLAPNGRDLIDECAS 180
 |||||
 121 LSGHMLPDCATVNSRTCAMINCOYSCDTEEGPQCLCPSSGLRLAPNGRDLIDECAS 180
 |||||
 181 GKVICPNRRCVNTFGSYCKCHIGFELQYISGRYDCIDINECTMDSHTCSHANCENTQ 240
 |||||
 181 GKVICPNRRCVNTFGSYCKCHIGFELQYISGRYDCIDINECTMDSHTCSHANCENTQ 240
 |||||
 241 GSPKCKCKQYKNGLRCSAIPENSVKVLRAPGTTIKRLAHKNSMKKAKIKNVT 300
 |||||
 241 GSPKCKCKQYKNGLRCSAIPENSVKVLRAPGTTIKRLAHKNSMKKAKIKNVT 300
 |||||
 301 PETRTPTPKVNIOPNRYEIVSRGNSHGKKGKNEKMEGLEDEKREKALKNDIEER 360
 |||||
 301 PETRTPTPKVNIOPNRYEIVSRGNSHGKKGKNEKMEGLEDEKREKALKNDIEER 360
 |||||
 361 SLRGDVFFPKVNEAGFGLLVORKALTSKLEHKDNLISVDCSFNHCICDWKQDREDDF 420
 |||||
 361 SLRGDVFFPKVNEAGFGLLVORKALTSKLEHKDNLISVDCSFNHCICDWKQDREDDF 420
 |||||
 421 WNPADRNAIGFYMAVPALAGHKDI GRLLKLLPDLQPSNFCLLPDYRLAGDKVGLRV 480
 |||||
 421 WNPADRNAIGFYMAVPALAGHKDI GRLLKLLPDLQPSNFCLLPDYRLAGDKVGLRV 480
 |||||
 481 FVKNNSNALAWKTTSEDEKWKTKIOLYQGTDTATKSIIFEAERGKGTGEIAVDGVLV 540
 |||||
 481 FVKNNSNALAWKTTSEDEKWKTKIOLYQGTDTATKSIIFEAERGKGTGEIAVDGVLV 540
 |||||
 541 SGLCPDLSLSVDD 553
 |||||
 541 SGLCPDLSLSVDD 553

```
FT Misc-difference 357 /label= Unknown
FT FT /note= "Xaa can be any amino acid"
FT FT 363..365
FT FT /note= "RGD motif"
FT FT 446..465
FT FT /note= "Transmembrane domain"
FT FT 509
FT FT /note= "Tyrosine phosphorylation site"
XX XX
XX US6392019-B1.
XX XX
XX 21-MAY-2002.
XX XX
XX 28-JUL-1999; 99US-00363316.
XX XX
XX 22-NOV-1997; 97US-00968800.
XX PR 12-FEB-1999; 99US-00249697.
XX XX
XX (FORD/) FORD J.
XX PA (YEUN/) YEUNG G.
XX XX
XX Ford J, Yeung G;
XX PI
XX XX
XX WPI: 2002-424836/45.
XX DR N-PSDB; AAD44332.
XX XX
XX Novel antibody specific for an epidermal growth factor repeat-containing
XX PT polypeptide, useful for the diagnosis of brain tumors, ulcers, leukemias,
XX PT and nervous disorders.
XX PS
XX Example 3; Fig 5; 92pp; English.
XX XX
XX The invention relates to an antibody specific for a 537 residue epidermal
XX CC growth factor (EGF) repeat-containing polypeptide sequence. The invention
XX CC is used for detecting the presence of EGF repeat containing polypeptides
XX CC in a sample, in the diagnosis of brain tumors, nervous disorders,
XX CC ulcers, and leukaemias. The present sequence is human EGF-repeat
XX CC containing protein
XX XX
XX Sequence 553 AA;
XX XX
Query Match 99.7%; Score 3050; DB 5; Length 553;
Best Local Similarity 99.6%; Pred No. 3.8e-212;
Matches 551; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 MPLPMSLALPLLSSWAGFGNAASARHGLLASARQPGVCHYGTKLACCCYGMRRNSKGV 60
DB 1 MPLPMSLALPLLPPWAGGFGNAASARHGLLASARQPGVCHYGTKLACCCYGMRRNSKGV 60
QY 61 CEATCEPCGKGEVGNKRCFPYGTGKTCQDVNEGCMKPRPCQHRVCVNTGSHYKCF 120
DB 61 CEATCEPCGKGEVGNKRCFPYGTGKTCQDVNEGCMKPRPCQHRVCVNTGSHYKCF 120
QY 121 LSGHMLMPDATCVNERTCAMINCOYSCDETBEGPQCLPSSGLRLAPNGRCLDIDECAS 180
DB 121 LSGHMLMPDATCVNERTCAMINCOYSCDETBEGPQCLPSSGLRLAPNGRCLDIDECAS 180
QY 181 GKVICPNRRCVNTFGSYCKCHIGFELQYISGRYDCIDINECTWDSHTCSHANCFTQ 240
DB 181 GKVICPNRRCVNTFGSYCKCHIGFELQYISGRYDCIDINECTWDSHTCSHANCFTQ 240
QY 241 GSFCKCKQGYKGNLGRCSAIPENSVKVLRAPGTIKDRIKKLAHKNMCKKAKIKNT 300
DB 241 GSFCKCKQGYKGNLGRCSAIPENSVKVLRAPGTIKDRIKKLAHKNMCKKAKIKNT 300
QY 301 PEPTRTPTPKVNLQPFNYEEIVSRGNSHGKKGNEEKKEGLEDEKREKALNDIEER 360
DB 301 PEPTRTPTPKVNLQPFNYEEIVSRGNSHGKKGNEEKKEGLEDEKREKALNDIEER 360
QY 361 SLRGDVFFPKVNEAGEFGLILVQRKALTSKLEHKDLNLSVDCSFNHHGICDMKQREDDDFD 420
DB 361 SLRGDVFFPKVNEAGEFGLILVQRKALTSKLEHKDLNLSVDCSFNHHGICDMKQREDDDFD 420
```

```
QY 421 WNPADRDNAIGFYMAVPALAGHKDIGRLKLLLPDLPQSNFCLLFDYRLAGDKVYKLRV 480
DB 421 WNPADRDNAIGFYMAVPALAGHKDIGRLKLLLPDLPQSNFCLLFDYRLAGDKVYKLRV 480
QY 481 FVKSNNALAWKEKTTSEDEKWKTKIQLYQGTDTAKSIIFAEERKGGTKGTGEIADVGVLLV 540
DB 481 FVKSNNALAWKEKTTSEDEKWKTKIQLYQGTDTAKSIIFAEERKGGTKGTGEIADVGVLLV 540
QY 541 SGLCPDLSLLSVDD 553
DB 541 SGLCPDLSLLSVDD 553
RESULT 13
ABG72935
ID ABG72935 standard; protein; 553 AA.
XX AC ABG72935;
XX DT 02-APR-2003 (first entry)
XX DE Novel human EGF-motif containing protein.
XX KW EGF; epidermal growth factor; cancer; lung cancer; brain cancer;
XX KW prostate cancer; breast cancer; skin cancer; lymphoma cancer;
XX KW sarcoma cancer; colon cancer; tumorigenicity; tumour site reduction;
XX KW cell proliferation inhibition; vaccine; antisense gene therapy; human.
XX OS Homo sapiens.
XX FT Key Location/Qualifiers
FT FT Misc-difference 357 /label= OTHER
FT FT /note= "Any amino acid"
XX XX
XX US2002132250-A1.
XX PD 19-SEP-2002.
XX XX
XX 15-OCT-2001; 2001US-00981649.
XX XX
XX 28-JUL-1999; 99US-00363316.
XX PR 13-OCT-2000; 2000US-00687860.
XX XX
XX (FORD/) FORD J E.
XX PA (YEUN/) YEUNG G.
XX PA (ZHOU/) ZHOU H.
XX PI Ford JB, Yeung G, Zhou H;
XX XX
XX WPI; 2003-174078/17.
XX DR N-PSDB; ABX14768.
XX XX
```

Detecting cancerous cells expressing polynucleotides/polypeptides in samples, by contacting samples with labeled polynucleotides complementary to polynucleotide or an antibody against the polypeptide and detecting complex formed.

Example 4; Fig 5; 78pp; English.

The invention describes a method of detecting a cancerous cell expressing a polynucleotide (I) or a polypeptide (II) in a biological sample, involving contacting the sample with a labelled polynucleotide complementary to (I) or an antibody or its fragment that specifically binds to (II), for a period sufficient to form a complex and detecting the complex, so that if a complex is detected, the cell is detected. The method is useful for detecting cancerous cell in a biological sample such as tissue, cell, blood, serum, lymphatic fluid, urine, and cerebrospinal fluid. The cancerous cell is from lung, brain, prostate, breast, skin, lymphoma, sarcoma and colon. Preferably the cancer cell is A549 cell, MCF-7 cell or SK-N-Mc cell. PC1 and PC2 are useful for inhibiting EGFL6 activity, inhibiting tumorigenicity, reducing tumour sites and inhibiting

CC proliferation of a cancer cell. This is the amino acid sequence of the
 CC novel human EGF (epidermal growth factor) motif containing protein
 XX
 SQ Sequence 553 AA;

Query Match 99.7%; Score 3050; DB 6; Length 553;
 Best Local Similarity 99.6%; Pred. No. 3.8e-212;
 Matches 551; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MPLPWSLALPLLSSWAGFGNNAASARHHGLLASARQPGVCHYGTKLACCYGWRNSKGV 60
 DB 1 MPLPWSLALPLLSPWAGFGNNAASARHHGLLASARQPGVCHYGTKLACCYGWRNSKGV 60

QY 61 CEATCEPGCKFGCECVGNPKRCRCPGYTGKTSQDVNECGMKPRPCQHRCVNTHSGYKFC 120
 DB 61 CEATCEPGCKFGCECVGNPKRCRCPGYTGKTSQDVNECGMKPRPCQHRCVNTHSGYKFC 120

QY 121 LSGHMLMPDATCVNSRTCAMINQYSCDTEBGPQCLCPSSGLRLAPNGRDCLDIDECAS 180
 DB 121 LSGHMLMPDATCVNSRTCAMINQYSCDTEBGPQCLCPSSGLRLAPNGRDCLDIDECAS 180

QY 181 GKVICPNRRCVNTFGSYCKCHIGPELQYISGRYDCIDINECTMDSHTCSHHANCFNTQ 240
 DB 181 GKVICPNRRCVNTFGSYCKCHIGPELQYISGRYDCIDINECTMDSHTCSHHANCFNTQ 240

QY 241 GSFCKCKQGYKNGLRCSAIPENSVEVLRAPGTIKDIRIKLLAHKNSMKKAKIKNVT 300
 DB 241 GSFCKCKQGYKNGLRCSAIPENSVEVLRAPGTIKDIRIKLLAHKNSMKKAKIKNVT 300

QY 301 PEPTRTPTPKVNLQPFNVEIIVSRGNSHGKKGNEKKEGLEDKREKALKNDIEER 360
 DB 301 PEPTRTPTPKVNLQPFNVEIIVSRGNSHGKKGNEKKEGLEDKREKALKNDIEER 360

QY 361 SLRGDVFVPKVNAGBFLILVORKALTSKLEHKLINISVDCSFNHCICDWKQDREDDFD 420
 DB 361 SLRGDVFVPKVNAGBFLILVORKALTSKLEHKLINISVDCSFNHCICDWKQDREDDFD 420

QY 421 WNPADRDNAIGFYMAVPALAGHKDIDGRLLKLLPDLPQPSNFCLLDYRLAGDKVKLRV 480
 DB 421 WNPADRDNAIGFYMAVPALAGHKDIDGRLLKLLPDLPQPSNFCLLDYRLAGDKVKLRV 480

QY 481 FVKNNSNALAWKTTSEDEKWKTKIQLYOGTDTATKSIIFEAERBGKGTGEIIVDGVLLV 540
 DB 481 FVKNNSNALAWKTTSEDEKWKTKIQLYOGTDTATKSIIFEAERBGKGTGEIIVDGVLLV 540

QY 541 SGLCPDLSLSVDD 553
 DB 541 SGLCPDLSLSVDD 553

RESULT 14
 ABU62258
 ID ABU62258 standard; protein; 553 AA.

XX
 AC ABU62258;

XX
 DT 01-SEP-2003 (first entry)

XX
 DE Epidermal growth factor motif protein EGFL6 #1.

XX Human; epidermal growth factor motif protein; EGFL6; cytostatic;
 KW neuroprotective; antibacterial; antiparasitic; antilipemic;
 KW antifertility; EGF-Agonist; EGF-Antagonist; cell growth; cancer;
 KW neurodegenerative disorder; leukaemia; brain tumour; lung tumour;
 KW breast tumour; gastrointestinal tumour; skin tumour; prostate tumour;
 KW carcinoma; parasite; biorhythm; fertility; metabolism; catabolism;
 KW anabolism.

XX Homo sapiens.

XX Key Location/Qualifiers
 FH Misc-difference 357

FT /label= OTHER

/note= "OTHER= any amino acid"

US2003036508-A1.

20-FEB-2003.

17-APR-2002; 2002US-00124986.

22-NOV-1997; 97US-00968800.

12-FEB-1999; 99US-00249697.

28-JUL-1999; 99US-00363316.

13-OCT-2000; 2000US-00687860.

15-OCT-2001; 2001US-00981649.

(FORD/) FORD J.

(YEUN/) YEUNG G.

(ZHOU/) ZHOU H.

Ford J, Yeung G, Zhou H;

WPI; 2003-492123/46.

N-PSDB; ACD25931.

Stimulating cell growth by contacting the cell with an EGFL6 polypeptide,

useful for the diagnosis and treatment of cancers and neurodegenerative

disorders.

Example 3; Fig 4; 86pp; English.

XX The invention describes a method of stimulating cell growth comprising
 CC contacting the cell with an EGFL6 polypeptide having at least 90 %
 CC sequence identity to a 553 amino acid sequence (SI), given in the
 CC specification, or its variant and/or fragment lacking a C-terminal
 CC portion of the EGFL6 polypeptide. The methods and compositions of the
 CC present invention are useful for the diagnosis and treatment of cancers
 CC and neurodegenerative disorders by stimulating cell growth. The cancers
 CC include leukaemia, brain, lung, breast, gastrointestinal, skin and
 CC prostate tumours and carcinomas. They can also be used in inhibiting the
 CC growth of infectious agents and parasites, effecting fertility, metabolism
 CC characteristic and biorhythms, effecting fertility, metabolism
 CC catabolism and anabolism of fats, vitamins, proteins and minerals, and
 CC effecting behavioural characteristics. This is the amino acid sequence of
 CC novel human epidermal growth factor motif protein EGFL6

XX Sequence 553 AA;

Query Match 99.7%; Score 3050; DB 6; Length 553;

Best Local Similarity 99.6%; Pred. No. 3.8e-212;

Matches 551; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MPLPWSLALPLLSSWAGFGNNAASARHHGLLASARQPGVCHYGTKLACCYGWRNSKGV 60

DB 1 MPLPWSLALPLLSPWAGFGNNAASARHHGLLASARQPGVCHYGTKLACCYGWRNSKGV 60

QY 61 CEATCEPGCKFGCECVGNPKRCRCPGYTGKTSQDVNECGMKPRPCQHRCVNTHSGYKFC 120

DB 61 CEATCEPGCKFGCECVGNPKRCRCPGYTGKTSQDVNECGMKPRPCQHRCVNTHSGYKFC 120

QY 121 LSGHMLMPDATCVNSRTCAMINQYSCDTEBGPQCLCPSSGLRLAPNGRDCLDIDECAS 180

DB 121 LSGHMLMPDATCVNSRTCAMINQYSCDTEBGPQCLCPSSGLRLAPNGRDCLDIDECAS 180

QY 181 GKVICPNRRCVNTFGSYCKCHIGPELQYISGRYDCIDINECTMDSHTCSHHANCFNTQ 240

DB 181 GKVICPNRRCVNTFGSYCKCHIGPELQYISGRYDCIDINECTMDSHTCSHHANCFNTQ 240

QY 241 GSFCKCKQGYKNGLRCSAIPENSVEVLRAPGTIKDIRIKLLAHKNSMKKAKIKNVT 300

DB 241 GSFCKCKQGYKNGLRCSAIPENSVEVLRAPGTIKDIRIKLLAHKNSMKKAKIKNVT 300

QY 301 PEPTRTPTPKVNLQPFNVEIIVSRGNSHGKKGNEKKEGLEDKREKALKNDIEER 360

DB 301 PEPTRTPTPKVNLQPFNVEIIVSRGNSHGKKGNEKKEGLEDKREKALKNDIEER 360

QY 361 SLRGDVFFPKVNEAGBGLILVORKALTSKLEHKLNIISVDCSFNHI CDWKQDREDDF 420
 Db 361 SLRGDVFFPKVNEAGBGLILVORKALTSKLEHKLNIISVDCSFNHI CDWKQDREDDF 420
 QY 421 WNPADRDNAIGFYMAVPALAGHKDIGRLKLLPDLQPOSNFCLLFDYRLAGDKVKGLRV 480
 Db 421 WNPADRDNAIGFYMAVPALAGHKDIGRLKLLPDLQPOSNFCLLFDYRLAGDKVKGLRV 480
 QY 481 FVXNSNNALAWEXTTSEDEKWKTKIQLYQGTDTATKSIIFEABERGKGTGEIADVGVLIV 540
 Db 481 FVXNSNNALAWEXTTSEDEKWKTKIQLYQGTDTATKSIIFEABERGKGTGEIADVGVLIV 540
 QY 541 SGLCPDSLLSVDD 553
 Db 541 SGLCPDSLLSVDD 553

RESULT 15
 AAB27224
 ID AAB27224 standard; protein; 554 AA.
 AC AAB27224;
 DT 27-MAR-2001 (first entry)
 DE Human EXMAD-2 SEQ ID NO: 2.
 KW Extracellular matrix and adhesion-associated protein; EXMAD; cancer;
 KW inflammation; reproductive disorder; cardiovascular disorder;
 KW immune disorder; musculoskeletal disorder; developmental disorder;
 KW gastrointestinal disorder; cell proliferation disorder.
 OS Homo sapiens.
 XX WO200068380-A2.
 XX 16-NOV-2000.
 PF 10-MAY-2000; 2000WO-US012811.
 PR 11-MAY-1999; 99US-0133643P.
 PR 23-AUG-1999; 99US-0150409P.
 XX (INCY-) INCYTE GENOMICS INC.
 XX Bandman O, Hillman JL, Tang YT, Lal P, Yue H, Baughn MR, Lu DAM;
 PI Azimzai Y;
 XX WPI; 2001-007395/01.
 DR N-PSDB; AAC66891.
 XX Isolated polynucleotide encoding extracellular matrix or adhesion-associated protein (EXMAD) useful for diagnosing, treating, or preventing disorders associated with expression of EXMAD such as proliferative, immune and genetic disorders.
 PS Claim 1; Page 88-89; 129pp; English.
 XX The present invention provides the protein and coding sequences for 25 novel extracellular matrix and adhesion-associated proteins (EXMADs). These are designated EXMAD-1, EXMAD-2, EXMAD-3, EXMAD-4, EXMAD-5, EXMAD-6, EXMAD-7, EXMAD-8, EXMAD-9, EXMAD-10, EXMAD-11, EXMAD-12, EXMAD-13, EXMAD-14, EXMAD-15, EXMAD-16, EXMAD-17, EXMAD-18, EXMAD-19, EXMAD-20, EXMAD-21, EXMAD-22, EXMAD-23, EXMAD-24 and EXMAD-25. They are useful in the prevention and treatment of cancers, cell proliferation, cardiovascular, reproductive, immune, musculoskeletal, developmental and gastrointestinal disorders and inflammation
 XX Sequence 554 AA;

Query Match 99.7%; Score 3049.5; DB 4; Length 554;
 Best Local Similarity 99.8%; Pred. No. 4.1e-212;

Matches 553; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
 QY 1 MELPWSLALPLLLSVVAGFGNAASARHHGLLLASARQPGVCHYGTKLACCYCWRRNSKGV 60
 Db 1 MELPWSLALPLLLSVVAGFGNAASARHHGLLLASARQPGVCHYGTKLACCYCWRRNSKGV 60
 QY 61 CEATCEPGCKFGEVGNPKRCFPYTGKTCSDQVNECMKPRPCQHRVNTHTGSKYKFC 120
 Db 61 CEATCEPGCKFGEVGNPKRCFPYTGKTCSDQVNECMKPRPCQHRVNTHTGSKYKFC 120
 QY 121 LSGHMLMPDATCVNRTCAMINCOYSCEDTEBGPCLCPSSGLRLAPNGRDCLDIDECAS 180
 Db 121 LSGHMLMPDATCVNRTCAMINCOYSCEDTEBGPCLCPSSGLRLAPNGRDCLDIDECAS 180
 QY 181 GKVICPNYRRCVNTFGSYCKHIGFELQYISGRYDCIDINECTMDSHTCSHHANCFNTQ 240
 Db 181 GKVICPNYRRCVNTFGSYCKHIGFELQYISGRYDCIDINECTMDSHTCSHHANCFNTQ 240
 QY 241 GSFCKCKQGYKGNGLRCSAIPENSVEVLRAPGTIKDRIKKLLAHKNSMKKKAKIKNVT 300
 Db 241 GSFCKCKQGYKGNGLRCSAIPENSVEVLRAPGTIKDRIKKLLAHKNSMKKKAKIKNVT 300
 QY 301 PEPTRTPTPKVNLQPFNYEEIVSRGNSHGKKGNEEKMEGLEDEKREKALKNDIEER 360
 Db 301 PEPTRTPTPKVNLQPFNYEEIVSRGNSHGKKGNEEKMEGLEDEKREKALKNDIEER 360
 QY 361 SLRGDVFFPKVNEAGBGLILVORKALTSKLEHK-DLNIISVDCSFNHI CDWKQDREDDF 419
 Db 361 SLRGDVFFPKVNEAGBGLILVORKALTSKLEHKADLNIISVDCSFNHI CDWKQDREDDF 420
 QY 420 WNPADRDNAIGFYMAVPALAGHKDIGRLKLLPDLQPOSNFCLLFDYRLAGDKVKGLR 479
 Db 421 WNPADRDNAIGFYMAVPALAGHKDIGRLKLLPDLQPOSNFCLLFDYRLAGDKVKGLR 480
 QY 480 FVXNSNNALAWEXTTSEDEKWKTKIQLYQGTDTATKSIIFEABERGKGTGEIADVGVLIV 539
 Db 481 FVXNSNNALAWEXTTSEDEKWKTKIQLYQGTDTATKSIIFEABERGKGTGEIADVGVLIV 540
 QY 540 VSGLCPSDLSLVDD 553
 Db 541 VSGLCPSDLSLVDD 554
 Search completed: May 21, 2004, 12:31:17
 Job time : 102.28 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 21, 2004, 12:29:27 ; Search time 32.1285 seconds
(without alignments)
888.592 Million cell updates/sec

Title: US-09-981-649A-24

Perfect score: 3060

Sequence: 1 MPLPWSLALPLLLSWVAGGF.....VDGVLVSLGICPDSLLSVD 553

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/prodata/2/iaa/5A COMB.pep.*
2: /cgn2_6/prodata/2/iaa/5B COMB.pep.*
3: /cgn2_6/prodata/2/iaa/6A COMB.pep.*
4: /cgn2_6/prodata/2/iaa/6B COMB.pep.*
5: /cgn2_6/prodata/2/iaa/PCUS COMB.pep.*
6: /cgn2_6/prodata/2/iaa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	DB ID	Description
1	3060	100.0	553	4	US-09-249-697A-19
2	3060	100.0	553	4	US-09-363-316B-24
3	3050	99.7	553	4	US-09-249-697A-6
4	3050	99.7	553	4	US-09-363-316B-6
5	2769	90.5	537	4	US-09-249-697A-4
6	2769	90.5	537	4	US-09-363-316B-4
7	2757	90.1	502	4	US-09-363-316B-18
8	1288	42.1	284	4	US-09-312-283C-389
9	1030	33.7	509	4	US-09-907-794A-315
10	1030	33.7	509	4	US-09-905-125A-315
11	1030	33.7	509	4	US-09-902-775A-315
12	911	29.8	164	4	US-09-249-697A-9
13	911	29.8	164	4	US-09-363-316B-9
14	602	19.7	100	4	US-09-249-697A-3
15	602	19.7	100	4	US-09-363-316B-3
16	354.5	11.6	1833	3	US-08-479-722B-2
17	354.5	11.6	1833	5	PCT-US95-02251-18
18	339	11.1	956	2	US-08-897-443-3
19	335.5	11.0	2703	1	US-08-185-432-19
20	335.5	11.0	2703	4	US-08-899-232-4
21	327.5	10.7	2321	4	US-09-230-652-2
22	323.5	10.6	915	4	US-09-907-794A-34
23	323.5	10.6	915	4	US-09-905-125A-34
24	323.5	10.6	915	4	US-09-902-775A-34
25	322.5	10.5	886	3	US-09-110-116-3
26	322.5	10.5	2556	1	US-08-083-590A-20
27	322.5	10.5	2556	3	US-08-532-384-20

28	320.5	10.5	638	2	US-08-997-443-1	Sequence 1, Appli
29	320.5	10.5	2556	1	US-08-185-432-17	Sequence 17, Appl
30	320.5	10.5	2556	4	US-08-899-232-2	Sequence 2, Appli
31	320	10.5	1394	6	5177197-30	Patent No. 5177197
32	314	10.3	448	4	US-09-409-096-4	Sequence 4, Appli
33	313	10.2	443	2	US-08-833-963C-2	Sequence 2, Appli
34	313	10.2	443	3	US-08-980-514-1	Sequence 1, Appli
35	313	10.2	448	2	US-08-884-072-1	Sequence 1, Appli
36	313	10.2	448	4	US-09-212-168-1	Sequence 1, Appli
37	311.5	10.2	2523	1	US-08-185-432-18	Sequence 18, Appl
38	311.5	10.2	2523	4	US-08-899-232-3	Sequence 3, Appli
39	311	10.2	835	4	US-09-284-819-6	Sequence 6, Appli
40	311	10.2	835	4	US-09-262-537-12	Sequence 12, Appl
41	311	10.2	1964	4	US-09-467-997-1	Sequence 1, Appli
42	310.5	10.1	810	2	US-08-820-170A-34	Sequence 34, Appl
43	310.5	10.1	810	3	US-09-055-699-34	Sequence 34, Appl
44	310.5	10.1	810	3	US-09-273-565-34	Sequence 34, Appl
45	310.5	10.1	810	4	US-09-565-538-34	Sequence 34, Appl

ALIGNMENTS

RESULT 1

US-09-249-697A-19
; Sequence 19, Application US/09249697A
; Patent No. 6392018
; GENERAL INFORMATION:
; APPLICANT: Ford, John
; TITLE OF INVENTION: NOVEL EGF MOTIF PROTEIN OBTAINED FROM A CDNA LIBRARY OF FETAL LIVER SPLEEN
; FILE REFERENCE: 24011-727
; CURRENT APPLICATION NUMBER: US/09/249,697A
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 08/968,800
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 19
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-249-697A-19

Query Match		100.0%;	Score 3060;	DB 4;	Length 553;
Best Local Similarity		100.0%;	Pred. No. 8.5e-242;		
Matches 553;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MPLPWSLALPLLLSWVAGGF	GNASARHHGLLASARQPGVCHYGTKLACCYGWRNSKGV	60	
DB	1	MPLPWSLALPLLLSWVAGGF	GNASARHHGLLASARQPGVCHYGTKLACCYGWRNSKGV	60	
QY	61	CEATCEPGCKGECVGNKRCFP	PGYTGKTCSDQVNECGMKRPPCQHRCVNTHGSKYKFC	120	
DB	61	CEATCEPGCKGECVGNKRCFP	PGYTGKTCSDQVNECGMKRPPCQHRCVNTHGSKYKFC	120	
QY	121	LSGHLMDATCVNSRTCA	MINCYSCDTEEGPCLCPSSGLRLAPNGRCLDIDECAS	180	
DB	121	LSGHLMDATCVNSRTCA	MINCYSCDTEEGPCLCPSSGLRLAPNGRCLDIDECAS	180	
QY	181	GKVICFYNNRCVNTFGSY	CKCHIGFELQYISGRYDCIDINECTMDSHTCSHANCFTQ	240	
DB	181	GKVICFYNNRCVNTFGSY	CKCHIGFELQYISGRYDCIDINECTMDSHTCSHANCFTQ	240	
QY	241	GSFKCKQYKNGNLRCSA	IPNSVKVLRAPGTIKDRIKKLLAHKNSMKKKAKIKNVT	300	
DB	241	GSFKCKQYKNGNLRCSA	IPNSVKVLRAPGTIKDRIKKLLAHKNSMKKKAKIKNVT	300	
QY	301	PEPRTPTPKYNLOPFN	YEEIVSRGNSHGKNGEKKGEGLEDEKKEEKALNDIER	360	
DB	301	PEPRTPTPKYNLOPFN	YEEIVSRGNSHGKNGEKKGEGLEDEKKEEKALNDIER	360	

```
QY 361 SLRGDVFPKVNAGEFGLILVQRKALTSKLEHKDLNISVDCSFNHHGICDWKQDREDDFD 420
Db 361 SLRGDVFPKVNAGEFGLILVQRKALTSKLEHKDLNISVDCSFNHHGICDWKQDREDDFD 420
QY 421 WNPADRNAIGFYMAVPALAGHKD1GRLLKLLPDLQPSNFCILFYRLAGDKVGLRV 480
Db 421 WNPADRNAIGFYMAVPALAGHKD1GRLLKLLPDLQPSNFCILFYRLAGDKVGLRV 480
QY 481 FVKNSNNALAWKTTSEDEKWKTKIQLYQGTATKSIIFEAERGKGTGEIAVDGVLV 540
Db 481 FVKNSNNALAWKTTSEDEKWKTKIQLYQGTATKSIIFEAERGKGTGEIAVDGVLV 540
QY 541 SGLCPDLSLLSVDD 553
Db 541 SGLCPDLSLLSVDD 553

RESULT 2
US-09-363-316B-24
; Sequence 24, Application US/09363316B
; Patent No. 6392019
; GENERAL INFORMATION:
; APPLICANT: Ford, John
; APPLICANT: Yeung, George
; TITLE OF INVENTION: EGF MOTIF PROTEIN MATERIALS AND METHODS
; FILE REFERENCE: 28110/35852
; CURRENT APPLICATION NUMBER: US/09/363.316B
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: US 09/249,697
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 08/968,800
; PRIOR FILING DATE: 1997-11-22
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 24
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Homo sapiens
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-363-316B-24

Query Match 100.0%; Score 3060; DB 4; Length 553;
Best Local Similarity 100.0%; Pred. No. 8.5e-242;
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPLPWSLALPLLWSWAGFGNAAARHHGLLASARQPGVCHYGTKLACCYGMWRNSKGV 60
Db 1 MPLPWSLALPLLWSWAGFGNAAARHHGLLASARQPGVCHYGTKLACCYGMWRNSKGV 60
QY 61 CEATCEPGCKFGCVGNKRCFCPGYTGKTCSDVNECGMKPRPCQHRVCNTHGSKYKFC 120
Db 61 CEATCEPGCKFGCVGNKRCFCPGYTGKTCSDVNECGMKPRPCQHRVCNTHGSKYKFC 120
QY 121 LSGHMLMPDATCVNSRTCAMINQYSCEDTEEGPQCLCPSSGLRLAPNGRDCLDIDECAS 180
Db 121 LSGHMLMPDATCVNSRTCAMINQYSCEDTEEGPQCLCPSSGLRLAPNGRDCLDIDECAS 180
QY 181 GKVICPNRRVCNVTFGSYCKCHIGFELQYISGRYDCIDINECTMDSHTCSHHANCNTQ 240
Db 181 GKVICPNRRVCNVTFGSYCKCHIGFELQYISGRYDCIDINECTMDSHTCSHHANCNTQ 240
QY 241 GSFCKCKQGYKGNGLRCSAIPENSVEVLAPGTIKDRIKLLAHKNSMKKKAKIKNVT 300
Db 241 GSFCKCKQGYKGNGLRCSAIPENSVEVLAPGTIKDRIKLLAHKNSMKKKAKIKNVT 300
QY 301 PEPTRTPTPKVNLQPNFYEEIVSRGNSHGKKGNEEKWKEGLEDEKREKALNDIEER 360
Db 301 PEPTRTPTPKVNLQPNFYEEIVSRGNSHGKKGNEEKWKEGLEDEKREKALNDIEER 360
QY 361 SLRGDVFPKVNAGEFGLILVQRKALTSKLEHKDLNISVDCSFNHHGICDWKQDREDDFD 420
Db 361 SLRGDVFPKVNAGEFGLILVQRKALTSKLEHKDLNISVDCSFNHHGICDWKQDREDDFD 420
QY 421 WNPADRNAIGFYMAVPALAGHKD1GRLLKLLPDLQPSNFCILFYRLAGDKVGLRV 480
```

```
Db 421 WNPADRNAIGFYMAVPALAGHKD1GRLLKLLPDLQPSNFCILFYRLAGDKVGLRV 480
QY 481 FVKNSNNALAWKTTSEDEKWKTKIQLYQGTATKSIIFEAERGKGTGEIAVDGVLV 540
Db 481 FVKNSNNALAWKTTSEDEKWKTKIQLYQGTATKSIIFEAERGKGTGEIAVDGVLV 540
QY 541 SGLCPDLSLLSVDD 553
Db 541 SGLCPDLSLLSVDD 553

RESULT 3
US-09-249-697A-6
; Sequence 6, Application US/09249697A
; Patent No. 6392018
; GENERAL INFORMATION:
; APPLICANT: Ford, John
; APPLICANT: Yeung, George
; TITLE OF INVENTION: NOVEL EGF MOTIF PROTEIN OBTAINED FROM A CDNA LIBRARY OF FETAL
; TITLE OF INVENTION: LIVER SPLEEN
; FILE REFERENCE: 24011-727
; CURRENT APPLICATION NUMBER: US/09/249,697A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 08/968,800
; PRIOR FILING DATE: 1997-11-22
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: VARIANT
; LOCATION: (1)...(553)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-249-697A-6

Query Match 99.7%; Score 3050; DB 4; Length 553;
Best Local Similarity 99.6%; Pred. No. 5.6e-241;
Matches 551; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MPLPWSLALPLLWSWAGFGNAAARHHGLLASARQPGVCHYGTKLACCYGMWRNSKGV 60
Db 1 MPLPWSLALPLLWSWAGFGNAAARHHGLLASARQPGVCHYGTKLACCYGMWRNSKGV 60
QY 61 CEATCEPGCKFGCVGNKRCFCPGYTGKTCSDVNECGMKPRPCQHRVCNTHGSKYKFC 120
Db 61 CEATCEPGCKFGCVGNKRCFCPGYTGKTCSDVNECGMKPRPCQHRVCNTHGSKYKFC 120
QY 121 LSGHMLMPDATCVNSRTCAMINQYSCEDTEEGPQCLCPSSGLRLAPNGRDCLDIDECAS 180
Db 121 LSGHMLMPDATCVNSRTCAMINQYSCEDTEEGPQCLCPSSGLRLAPNGRDCLDIDECAS 180
QY 181 GKVICPNRRVCNVTFGSYCKCHIGFELQYISGRYDCIDINECTMDSHTCSHHANCNTQ 240
Db 181 GKVICPNRRVCNVTFGSYCKCHIGFELQYISGRYDCIDINECTMDSHTCSHHANCNTQ 240
QY 241 GSFCKCKQGYKGNGLRCSAIPENSVEVLAPGTIKDRIKLLAHKNSMKKKAKIKNVT 300
Db 241 GSFCKCKQGYKGNGLRCSAIPENSVEVLAPGTIKDRIKLLAHKNSMKKKAKIKNVT 300
QY 301 PEPTRTPTPKVNLQPNFYEEIVSRGNSHGKKGNEEKWKEGLEDEKREKALNDIEER 360
Db 301 PEPTRTPTPKVNLQPNFYEEIVSRGNSHGKKGNEEKWKEGLEDEKREKALNDIEER 360
QY 361 SLRGDVFPKVNAGEFGLILVQRKALTSKLEHKDLNISVDCSFNHHGICDWKQDREDDFD 420
Db 361 SLRGDVFPKVNAGEFGLILVQRKALTSKLEHKDLNISVDCSFNHHGICDWKQDREDDFD 420
QY 421 WNPADRNAIGFYMAVPALAGHKD1GRLLKLLPDLQPSNFCILFYRLAGDKVGLRV 480
Db 421 WNPADRNAIGFYMAVPALAGHKD1GRLLKLLPDLQPSNFCILFYRLAGDKVGLRV 480
```

QY 481 FVXNSNNALAWKTTSEDEKWKTKGIQLYQGTDTATKSIIFAEARGKGTGBIADVGVLLV 540
 Db |||||
 QY 481 FVXNSNNALAWKTTSEDEKWKTKGIQLYQGTDTATKSIIFAEARGKGTGBIADVGVLLV 540
 Db |||||

RESULT 4
 US-09-363-316B-6
 ; Sequence 6, Application US/09363316B
 ; Patent No. 6392019
 ; GENERAL INFORMATION:
 ; APPLICANT: Ford, John
 ; APPLICANT: Yeung, George
 ; TITLE OF INVENTION: EGF MOTIF PROTEIN MATERIALS AND METHODS
 ; FILE REFERENCE: 28110735852
 ; CURRENT APPLICATION NUMBER: US/09/363,316B
 ; CURRENT FILING DATE: 1999-07-28
 ; PRIOR APPLICATION NUMBER: US 09/249,697
 ; PRIOR FILING DATE: 1999-02-12
 ; PRIOR APPLICATION NUMBER: US 08/968,800
 ; PRIOR FILING DATE: 1997-11-22
 ; NUMBER OF SEQ ID NOS: 24
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 6
 ; LENGTH: 553
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (357)
 ; OTHER INFORMATION: Xaa = any amino acid
 US-09-363-316B-6

Query Match 99.7%; Score 3050; DB 4; Length 553;
 Best Local Similarity 99.6%; Pred. No. 5.6e-241;
 Matches 551; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MFLPWSLALPLLWSVAGGFNGASARHGLLASARQGVCHYGTKLACCGWERNKGV 60
 Db 1 MFLPWSLALPLLWPVAGGFNGASARHGLLASARQGVCHYGTKLACCGWERNKGV 60
 QY 61 CEATCEPGCKEFCVGNPKRCFFPGYTGKTCSDVNECGMKPRPCQHRNVTHGSYKFC 120
 Db 61 CEATCEPGCKEFCVGNPKRCFFPGYTGKTCSDVNECGMKPRPCQHRNVTHGSYKFC 120
 QY 121 LSGHMLMPDATCVNSRTCAMINCOYSCEDTEEGPQCLCPSSGLRLAPNRCGLDIDECAS 180
 Db 121 LSGHMLMPDATCVNSRTCAMINCOYSCEDTEEGPQCLCPSSGLRLAPNRCGLDIDECAS 180
 QY 181 GKVICPNVRRCVNTFGSYCKCHIGFELQYISGRYDCIDINECTMDSHTCSHHANCNTQ 240
 Db 181 GKVICPNVRRCVNTFGSYCKCHIGFELQYISGRYDCIDINECTMDSHTCSHHANCNTQ 240
 QY 241 GSFCKCKQYKGNGLRCSAIPENSVEVLAPGTIKDRIKKLAHKNMKKAKIKNVT 300
 Db 241 GSFCKCKQYKGNGLRCSAIPENSVEVLAPGTIKDRIKKLAHKNMKKAKIKNVT 300
 QY 301 PEPTPTPKVNLQPFNYEETVSRGNSHGKGNKEEKKEGLEDEKREKALKNDEER 360
 Db 301 PEPTPTPKVNLQPFNYEETVSRGNSHGKGNKEEKKEGLEDEKREKALKNDEER 360
 QY 361 SLRGDVPFPKNEAGEFGLIIVORKALTSKLEHKLNTISVDCSFNHCDCWKQDREDDFD 420
 Db 361 SLRGDVPFPKNEAGEFGLIIVORKALTSKLEHKLNTISVDCSFNHCDCWKQDREDDFD 420
 QY 421 WNPADRONAIGFYMAVPALAGHKDKIGRLKLLLPDLPQSNFCLLDYRLAGDKVGLRV 480
 Db 421 WNPADRONAIGFYMAVPALAGHKDKIGRLKLLLPDLPQSNFCLLDYRLAGDKVGLRV 480

QY 481 FVXNSNNALAWKTTSEDEKWKTKGIQLYQGTDTATKSIIFAEARGKGTGBIADVGVLLV 540
 Db |||||
 QY 481 FVXNSNNALAWKTTSEDEKWKTKGIQLYQGTDTATKSIIFAEARGKGTGBIADVGVLLV 540
 Db |||||

QY 541 SGLCPDLSLLSVDD 553
 Db |||||

RESULT 5
 US-09-249-697A-4
 ; Sequence 4, Application US/09249697A
 ; Patent No. 6392018
 ; GENERAL INFORMATION:
 ; APPLICANT: Ford, John
 ; APPLICANT: Yeung, George
 ; TITLE OF INVENTION: NOVEL EGF MOTIF PROTEIN OBTAINED FROM A CDNA LIBRARY OF FETAL
 ; TITLE OF INVENTION: LIVER SPLEEN
 ; FILE REFERENCE: 24011-727
 ; CURRENT APPLICATION NUMBER: US/09/249,697A
 ; CURRENT FILING DATE: 1999-02-12
 ; PRIOR APPLICATION NUMBER: US 08/968,800
 ; PRIOR FILING DATE: 1997-11-22
 ; NUMBER OF SEQ ID NOS: 19
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 4
 ; LENGTH: 537
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: VARIANT
 ; LOCATION: (1)...(537)
 ; OTHER INFORMATION: Xaa = Any Amino Acid
 US-09-249-697A-4

Query Match 90.5%; Score 2769; DB 4; Length 537;
 Best Local Similarity 99.8%; Pred. No. 4.9e-218;
 Matches 501; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 52 GWRNSKGVCEANCPGCKEFCVGNPKRCFFPGYTGKTCSDVNECGMKPRPCQHRNV 111
 Db 1 GWRNSKGVCEANCPGCKEFCVGNPKRCFFPGYTGKTCSDVNECGMKPRPCQHRNV 60
 QY 112 THGSYKFCCLSGHMLMPDATCVNSRTCAMINCOYSCEDTEEGPQCLCPSSGLRLAPNRCGLDIDECAS 171
 Db 61 THGSYKFCCLSGHMLMPDATCVNSRTCAMINCOYSCEDTEEGPQCLCPSSGLRLAPNRCGLDIDECAS 120
 QY 172 CLDIDECASGVICPNVRRCVNTFGSYCKCHIGFELQYISGRYDCIDINECTMDSHTCS 231
 Db 121 CLDIDECASGVICPNVRRCVNTFGSYCKCHIGFELQYISGRYDCIDINECTMDSHTCS 180
 QY 232 HHANCFNTQSGFKCKQYKGNGLRCSAIPENSVEVLAPGTIKDRIKKLAHKNMK 291
 Db 181 HHANCFNTQSGFKCKQYKGNGLRCSAIPENSVEVLAPGTIKDRIKKLAHKNMK 240
 QY 292 KXAKINVTPEPTPTPKVNLQPFNYEETVSRGNSHGKGNKEEKKEGLEDEKREK 351
 Db 241 KXAKINVTPEPTPTPKVNLQPFNYEETVSRGNSHGKGNKEEKKEGLEDEKREK 300
 QY 352 ALKNDIEERSLGRDVPFPKNEAGEFGLIIVORKALTSKLEHKLNTISVDCSFNHCDCWK 411
 Db 301 ALKNDIEERSLGRDVPFPKNEAGEFGLIIVORKALTSKLEHKLNTISVDCSFNHCDCWK 360
 QY 412 KQDREDDFDWNPADRONAIGFYMAVPALAGHKDKIGRLKLLLPDLPQSNFCLLDYRLA 471
 Db 361 KQDREDDFDWNPADRONAIGFYMAVPALAGHKDKIGRLKLLLPDLPQSNFCLLDYRLA 420
 QY 472 GDKVGLRVFVXNSNNALAWKTTSEDEKWKTKGIQLYQGTDTATKSIIFAEARGKGTGE 531
 Db 421 GDKVGLRVFVXNSNNALAWKTTSEDEKWKTKGIQLYQGTDTATKSIIFAEARGKGTGE 480
 QY 532 IADVGVLLVSGLCPLDSLLSVDD 553
 Db |||||

Db 481 IAVDGVLLVSLGCLPDSLLSVDD 502

RESULT 6

US-09-363-316B-4
; Sequence 4, Application US/09363316B
; Patent No. 6392019
; GENERAL INFORMATION:
; APPLICANT: Ford, John
; APPLICANT: Yeung, George
; TITLE OF INVENTION: EGF MOTIF PROTEIN MATERIALS AND METHODS
; FILE REFERENCE: 28110/35852
; CURRENT FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: US 09/249,697
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 08/968,800
; PRIOR FILING DATE: 1997-11-22
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 537
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc_feature
; LOCATION: (503)
; OTHER INFORMATION: Xaa = any amino acid
US-09-363-316B-4

Query Match 90.5%; Score 2769; DB 4; Length 537;

Best Local Similarity 99.8%; Pred. No. 4.9e-218; Mismatches 1; Indels 0; Gaps 0;

Qy	52	GWRRNSKGVCEATCPGCKFGECVGNKRCFPFGYTGKTCSDVNECGMKPRPCOHRCVN	111
Db	1	GWRRNSKGVCEATCPGCKFGECVGNKRCFPFGYTGKTCSDVNECGMKPRPCOHRCVN	60
Qy	112	THGSYKCFCLSGHMLMPDATCVNSRTCAMINQVSCDETEEGPQCLCPSSGLRLAPNGRD	171
Db	61	THGSYKCFCLSGHMLMPDATCVNSRTCAMINQVSCDETEEGPQCLCPSSGLRLAPNGRD	120
Qy	172	CLDIDECASGKVICPNRRCVNTFGSYCKCHIGFELQYISGRYDCIDINECTWDSHTCS	231
Db	121	CLDIDECASGKVICPNRRCVNTFGSYCKCHIGFELQYISGRYDCIDINECTWDSHTCS	180
Qy	232	HHANCFNTQGSFKCKQGYKGNLRCSAIPENSVKVLRAPGTIKDRIKLLAHKNSMK	291
Db	181	HHANCFNTQGSFKCKQGYKGNLRCSAIPENSVKVLRAPGTIKDRIKLLAHKNSMK	240
Qy	292	KKAKIKNVTPEPTRTPPKVNLQPFNYEEIVSRGNSHGKKGNEEKKEGLEDEKREEK	351
Db	241	KKAKIKNVTPEPTRTPPKVNLQPFNYEEIVSRGNSHGKKGNEEKKEGLEDEKREEK	300
Qy	352	ALKNDIEERSLRGVDVFPKVNAGEFGLILVQRKALTSLKLEHKLNIISVDCSFNHGICDW	411
Db	301	ALKNDIEERSLRGVDVFPKVNAGEFGLILVQRKALTSLKLEHKLNIISVDCSFNHGICDW	360
Qy	412	KQREDDFDWNADRDNAIGFYMAVPALAGHKDIOIGRLKLLPDLQPOSNFCFLFDYRLA	471
Db	361	KQREDDFDWNADRDNAIGFYMAVPALAGHKDIOIGRLKLLPDLQPOSNFCFLFDYRLA	420
Qy	472	GDKVGLRVFVKNSNNALAWKTTSEDEKWKTKIQLYQGTDTATKSIIFEAERGKGTGE	531
Db	421	GDKVGLRVFVKNSNNALAWKTTSEDEKWKTKIQLYQGTDTATKSIIFEAERGKGTGE	480
Qy	532	IAVDGVLLVSLGCLPDSLLSVDD 553	
Db	481	IAVDGVLLVSLGCLPDSLLSVDD 502	

RESULT 7

US-09-363-316B-18

; Sequence 18, Application US/09363316B
; Patent No. 6392019
; GENERAL INFORMATION:
; APPLICANT: Ford, John
; APPLICANT: Yeung, George
; TITLE OF INVENTION: EGF MOTIF PROTEIN MATERIALS AND METHODS
; FILE REFERENCE: 28110/35852
; CURRENT FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: US 09/249,697
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 08/968,800
; PRIOR FILING DATE: 1997-11-22
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 18
; LENGTH: 502
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc_feature
; LOCATION: (501-502)
; OTHER INFORMATION: Xaa = any amino acid
US-09-363-316B-18

Query Match 90.1%; Score 2757; DB 4; Length 502;

Best Local Similarity 99.8%; Pred. No. 4.3e-217; Mismatches 1; Indels 0; Gaps 0;

Qy	52	GWRRNSKGVCEATCPGCKFGECVGNKRCFPFGYTGKTCSDVNECGMKPRPCOHRCVN	111
Db	1	GWRRNSKGVCEATCPGCKFGECVGNKRCFPFGYTGKTCSDVNECGMKPRPCOHRCVN	60
Qy	112	THGSYKCFCLSGHMLMPDATCVNSRTCAMINQVSCDETEEGPQCLCPSSGLRLAPNGRD	171
Db	61	THGSYKCFCLSGHMLMPDATCVNSRTCAMINQVSCDETEEGPQCLCPSSGLRLAPNGRD	120
Qy	172	CLDIDECASGKVICPNRRCVNTFGSYCKCHIGFELQYISGRYDCIDINECTWDSHTCS	231
Db	121	CLDIDECASGKVICPNRRCVNTFGSYCKCHIGFELQYISGRYDCIDINECTWDSHTCS	180
Qy	232	HHANCFNTQGSFKCKQGYKGNLRCSAIPENSVKVLRAPGTIKDRIKLLAHKNSMK	291
Db	181	HHANCFNTQGSFKCKQGYKGNLRCSAIPENSVKVLRAPGTIKDRIKLLAHKNSMK	240
Qy	292	KKAKIKNVTPEPTRTPPKVNLQPFNYEEIVSRGNSHGKKGNEEKKEGLEDEKREEK	351
Db	241	KKAKIKNVTPEPTRTPPKVNLQPFNYEEIVSRGNSHGKKGNEEKKEGLEDEKREEK	300
Qy	352	ALKNDIEERSLRGVDVFPKVNAGEFGLILVQRKALTSLKLEHKLNIISVDCSFNHGICDW	411
Db	301	ALKNDIEERSLRGVDVFPKVNAGEFGLILVQRKALTSLKLEHKLNIISVDCSFNHGICDW	360
Qy	412	KQREDDFDWNADRDNAIGFYMAVPALAGHKDIOIGRLKLLPDLQPOSNFCFLFDYRLA	471
Db	361	KQREDDFDWNADRDNAIGFYMAVPALAGHKDIOIGRLKLLPDLQPOSNFCFLFDYRLA	420
Qy	472	GDKVGLRVFVKNSNNALAWKTTSEDEKWKTKIQLYQGTDTATKSIIFEAERGKGTGE	531
Db	421	GDKVGLRVFVKNSNNALAWKTTSEDEKWKTKIQLYQGTDTATKSIIFEAERGKGTGE	480
Qy	532	IAVDGVLLVSLGCLPDSLLSV 551	
Db	481	IAVDGVLLVSLGCLPDSLLSV 500	

RESULT 8

US-09-312-283C-389
; Sequence 389, Application US/09312283C
; Patent No. 6573095
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna

```
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James G.
; APPLICANT: Kumble, Krihanand D.
; TITLE OF INVENTION: Compositions Isolated from Skin Cells
; TITLE OF INVENTION: and Methods for Their Use
; FILE REFERENCE: 11000.1011c2
; CURRENT APPLICATION NUMBER: US/09/312,283C
; CURRENT FILING DATE: 1999-05-14
; NUMBER OF SEQ ID NOS: 425
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 389
; LENGTH: 284
; TYPE: PRT
; ORGANISM: Mouse
US-09-312-283C-389

Query Match 42.1%; Score 1288; DB 4; Length 284;
Best Local Similarity 82.6%; Pred. No. 2.2e-97;
Matches 219; Conservative 14; Mismatches 30; Indels 2; Gaps 1;

QY 1 MFLPWSLALPLLISWVAGGFGNAASRHHGLLASAROPGVHYGTKLACCVGMRNSKGV 60
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 61 CEATCEPGCKFGECVGNPKRCPCPGYTGKTCSDVNECGMKPRPCQHRVCVNTHGSYKFC 120
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 81 CEAVCEPRCKFGECVGNPKRCPCPGYTGKTCSDVNECAFKPRPCQHRVCVNTHGSYKFC 140
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 121 LSGHMLPMDATCVNSRTCAMINCOYSCDETEBGPQCLCPSSGRLAPNGRDLIDIECAS 180
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 141 LSGHMLPMDATCVNSRTCAMINCOYSCDETEBGPQCLCPSSGRLAPNGRDLIDIECAS 200
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 181 GKVICPVNRVCVNTFGSYCKCHIGFPELOYSIGRYDCIDINECTMDSHTCSHANCFTQ 240
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 201 SKAVCSNRVCVNTFGSYCKCHIGFELKVIISRYDCVDINECTNTRTCSPRANCLNTQ 260
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 241 GSPKCKCKQYKGNGLRCSAIPEN 264
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 9
US-09-907-794A-315
; Sequence 315, Application US/09907794A
; Patent No. 6635468
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnovers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kijavlin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
```

```
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/907,794A
; CURRENT FILING DATE: 2001-07-17
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 315
; LENGTH: 509
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-907-794A-315

Query Match 33.7%; Score 1030; DB 4; Length 509;
Best Local Similarity 38.3%; Pred. No. 5.6e-76;
Matches 199; Conservative 84; Mismatches 141; Indels 96; Gaps 13;

QY 39 GVCHYGTKLACCYGWRNRNSKGYCEATCEPGCKFGECVGNPKRCPCPGYTGKTCSDVNEC 98
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 99 GMPKPRCQHRVCVNTHGSYKFCISLGHMLPMDATCVNSRTCAMINCOYSCDETEBGPQCL 158
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 94 GLKPRCKRHCNMTYGSYKCYCLNGTLMPLDGSCLSSALTCSMANCOYGDVVVGQIRQC 153
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 159 PSSGLRLAPNGRDLIDIECASGKVICPVNRVCVNTFGSYCKCHIGFPELOYSIGRYDCI 218
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 154 PFGHLAPDGRITCVDDIECATGRASCPFRVCVNTFGSYCKCHIGFPELOYSIGRYDCI 213
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 219 DINECTMDSHTCSHANCFTQSGFKCKCKQYKGNGLRCSAIPENSVK-----EVLRAP 273
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 214 DIDECSLGQYQCSFARCVNVRGSKCKEYQDGLTCTVYIPKVMIEPSGPIHVPGKN 273
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 274 GTI-----KDKIKLLAHNSMKKAKIKNVTPEPTRTPP--- 309
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 274 GTILKGDGTGNNNWIPDVGS*WMPKTPYIPITINRPTSKPTR---PTPKPTPTPTPP 330
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 310 -----KVNLOPFNYEE-----IVSRGNSHGKKGNEEKMEKGLGDEKREKALKN 355
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 331 PPPLTELRTPPTPTPTPTPTPTGLTTIAPASTPPG-----GITVDNRVQ----- 375
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```



```

; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/902,775A
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 315
; LENGTH: 509
; TYPE: PRT
; ORGANISM: Homo Sapien
; US-09-902-775A-315

Query Match      33.7%; Score 1030; DB 4; Length 509;
Best Local Similarity 38.3%; Pred. No. 5.6e-76;
Matches 199; Conservative 84; Mismatches 141; Indels 96; Gaps 13;

QY 39 GVCHYGTKLACCYGWRNRNSKVCCEATCEPGCKFGCECVGNPKRCRCFFPGYTGKTCSDVNEC 98
Db 34 GLCFYGGRIDCCWGWARQSGQCPVCPQRCRKGECIGPNKCKCHPGYAGKTCNQDLNEC 93

QY 99 GMPKRPQHRVNTGYSKYKFCFLSGHMLMPDTCVNSRTCAVNSQVSCDTEBGPQCLC 158
Db 94 GLKPRPKCHRMNTYGSKYKCYCLNGMLMPDTCVNSQVSCDTEBGPQCLC 153

QY 159 PSSGLRLAPNGRDLCDIDECASGVNRCVNTFGSYCKCHIGFELQYISGRYDCI 218

```

```

Db 154 PSPGLHLPDGRCTCDVDDECATGRASCPFRFRQCVNTFGSYCKCHKGFDLMYIGKYQCH 213
QY 219 DIRECTMDSHTCSHANCNTQGSFKCKCKGKYGKNGLRCSAIPENSVK-----EVLRA 273
Db 214 DIDECSLGQYQCSFARCYNVGRSYKCKRGGYQDGLTCVYIPKWMIEPSGPIHVPKGN 273
QY 274 GTI-----KDIRIKLLAHKNSMKKAKIKNVTPPTPTPTP-- 309
Db 274 GTILKGDGTNNWIPDVGSTWMPKTEVPIPIINRPTSKPTTR---PTPKPTPIPTPPP 330
QY 310 -----KVNLOPFNYEE-----IVSRGNSGGKKGNEKKMEGLEDEKREKALKN 355
Db 331 PPPLPTBLRPLPPTPTPTPTPTTGLTTTAPAASTPPG-----GITVDNRVQ----- 375
QY 356 DIERSLRGVDVFPKVNAGEFGLILVQRKALTSKLEKDLNISVDCSFNHGICDWMKODR 415
Db 376 -TPQKPRGDF-----SVLVH-----SCNFHGLCGWIREK 406
QY 416 EDDFDNPNADRNAIGFYMAVPALAGHKDIGRLKLLPDLQPOSNFCLLFDYELAGDKV 475
Db 407 DNDLHWEPI-RDPAGGYLTVSAKAPGKAAARLVPLGLRMLHSGDLCLSFRHKVTLGHS 465
QY 476 GKLRVFK--NSNNALAWKTTSEDEKWKTKQLYQGTD 513
Db 466 GTLQVFRVKGHAHAALWCRNGHG--WRQTQITL-RGAD 502

RESULT 12
US-09-249-697A-9
; Sequence 9, Application US/09249697A
; Patent No. 6392018
; GENERAL INFORMATION:
; APPLICANT: Ford, John
; APPLICANT: Yeung, George
; TITLE OF INVENTION: NOVEL EGF MOTIF PROTEIN OBTAINED FROM A CDNA LIBRARY OF FETAL
; TITLE OF INVENTION: LIVER SPLEEN
; FILE REFERENCE: 24011-727
; CURRENT APPLICATION NUMBER: US/09/249,697A
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 08/966,800
; PRIOR FILING DATE: 1997-11-22
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 164
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-249-697A-9

Query Match      29.8%; Score 911; DB 4; Length 164;
Best Local Similarity 91.1%; Pred. No. 6.6e-67;
Matches 164; Conservative 0; Mismatches 0; Indels 16; Gaps 5;

QY 80 CRCFCYTGKTCSDQVNECGMKPRPCQHRVCVNTGYSKYKFCFLSGHMLMPDTCVNSRTCA 139
Db 1 CRCFCYTGKTCSQ-VNECGMKPRPCQHRVCVNTGYSKYKFCFLSGHMLMPD---VNSRTCA 56

QY 140 MINQYSCDTEBGPQCLCPSSGLRLAPNGRDLCDIDECASGVNRCVNTFGSY 199
Db 57 MINQYSCDTEBGPQCLCPSSGLRLAPN-----IDECASGVNRCVNTFGSY 110

QY 200 CKCHIGFELQYISGRYDCIDINECTMDSHTCSHANCNTQGSFKCKCKGKNGLRCS 259
Db 111 CKCHIGFELQYISGR-----INECTMDSHTCSHANCNTQGSF-CKCKQYKNGLRCS 164

RESULT 13
US-09-363-316B-9
; Sequence 9, Application US/09363316B
; Patent No. 6392019
; GENERAL INFORMATION:
; APPLICANT: Ford, John

```

```
; APPLICANT: Yeung, George
; TITLE OF INVENTION: EGF MOTIF PROTEIN MATERIALS AND METHODS
; FILE REFERENCE: 28110/35852
; CURRENT APPLICATION NUMBER: US/09/363,316B
; CURRENT FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: US 09/249,697
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 08/968,800
; PRIOR FILING DATE: 1997-11-22
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 164
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-363-316B-9

Query Match      29.8%; Score 911; DB 4; Length 164;
Best Local Similarity 91.1%; Pred. No. 6.1e-67;
Matches 164; Conservative 0; Mismatches 0; Indels 16; Gaps 5;

QY 80 CRCFGYGTCTCSQDVNECGMKPRPCQHRVNTGSHYKFCFLSGHMLMPDATCVNSRTCA 139
DB 1 CRCFGYGTCTCSQ-VNECGMKPRPCQHRVNTGSHYKFCFLSGHMLMPD---VNSRTCA 56

QY 140 MINQYSCEDTEEGPQCLCPSSGLRLAPNDRDCLDIDECASGKVICPYNRRVCVNTFGSY 199
DB 57 MINQYSCEDTEEGPQCLCPSSGLRLAPN-----IDECASGKVICPYNRRVCVNTFGSY 110

QY 200 CKCHIGFELQYISGRYDCIDINECTMDSHTCSHHANCFNTQGSFKCKQYKGNGLRCS 259
DB 111 CKCHIGFELQYISGR-----INECTMDSHTCSHHANCFNTQGSF-CKCKQYKGNGLRCS 164

RESULT 14
US-09-249-697A-3
; Sequence 3, Application US/09249697A
; Patent No. 6392018
; GENERAL INFORMATION:
; APPLICANT: Yeung, George
; TITLE OF INVENTION: NOVEL EGF MOTIF PROTEIN OBTAINED FROM A CDNA LIBRARY OF FETAL
; FILE REFERENCE: 24011-727
; CURRENT APPLICATION NUMBER: US/09/249,697A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 08/968,800
; PRIOR FILING DATE: 1997-11-22
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 100
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-249-697A-3

Query Match      19.7%; Score 602; DB 4; Length 100;
Best Local Similarity 100.0%; Pred. No. 6.1e-42;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 GWRNSKGVCEATCEPGCKFGECVGNPKRCFPGYTGKTCSDVNECGMKPRPCQHRCVN 111
DB 1 GWRNSKGVCEATCEPGCKFGECVGNPKRCFPGYTGKTCSDVNECGMKPRPCQHRCVN 60

QY 112 THGSYKFCFLSGHMLMPDATCVNSRTCAMINCOYSCEDTE 151
DB 61 THGSYKFCFLSGHMLMPDATCVNSRTCAMINCOYSCEDTE 100

RESULT 15
US-09-363-316B-3
; Sequence 3, Application US/09363316B
; Patent No. 6392019
```

```
; GENERAL INFORMATION:
; APPLICANT: Ford, John
; TITLE OF INVENTION: EGF MOTIF PROTEIN MATERIALS AND METHODS
; FILE REFERENCE: 28110/35852
; CURRENT APPLICATION NUMBER: US/09/363,316B
; CURRENT FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: US 09/249,697
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 08/968,800
; PRIOR FILING DATE: 1997-11-22
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 100
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-363-316B-3

Query Match      19.7%; Score 602; DB 4; Length 100;
Best Local Similarity 100.0%; Pred. No. 6.1e-42;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 GWRNSKGVCEATCEPGCKFGECVGNPKRCFPGYTGKTCSDVNECGMKPRPCQHRCVN 111
DB 1 GWRNSKGVCEATCEPGCKFGECVGNPKRCFPGYTGKTCSDVNECGMKPRPCQHRCVN 60

QY 112 THGSYKFCFLSGHMLMPDATCVNSRTCAMINCOYSCEDTE 151
DB 61 THGSYKFCFLSGHMLMPDATCVNSRTCAMINCOYSCEDTE 100
```

Search completed: May 21, 2004, 12:34:41
Job time : 33.1285 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 21, 2004, 12:31:58 ; Search time 75.9401 Seconds
(without alignments)
2031.122 Million cell updates/sec

Title: US-09-981-649A-24

Perfect score: 3060
Sequence: 1 MPLPWSLALPLLSSWAGGF.....VDGVLLVSLGLCPDLSLLSVD 553

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1149313 seqs, 278921704 residues

Total number of hits satisfying chosen parameters: 1149313

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA.*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3060	100.0	553	9	US-09-981-649A-24
2	3060	100.0	553	10	US-09-796-753-14
3	3060	100.0	553	12	US-10-058-270A-102
4	3060	100.0	553	12	US-10-399-123-24
5	3060	100.0	553	14	US-10-124-986-24
6	3060	100.0	553	14	US-10-136-227A-24
7	3060	100.0	553	14	US-10-112-881-24
8	3060	100.0	553	15	US-10-295-027-494
9	3060	100.0	553	15	US-10-295-027-812
10	3060	100.0	553	15	US-10-295-027-841
11	3060	100.0	553	15	US-10-173-999-46
12	3060	100.0	553	16	US-10-188-832-189
13	3050	99.7	553	9	US-09-981-649A-6
14	3050	99.7	553	12	US-10-399-123-6
15	3050	99.7	553	14	US-10-124-986-6

16	3050	99.7	553	14	US-10-136-227A-6	Sequence 6, Appli
17	3050	99.7	553	14	US-10-112-881-6	Sequence 6, Appli
18	3049.5	99.7	554	9	US-09-981-649A-30	Sequence 30, Appl
19	3049.5	99.7	554	9	US-09-981-649A-32	Sequence 32, Appl
20	3049.5	99.7	554	12	US-10-399-123-30	Sequence 30, Appl
21	3049.5	99.7	554	12	US-10-399-123-32	Sequence 32, Appl
22	3049.5	99.7	554	14	US-10-124-986-30	Sequence 30, Appl
23	3049.5	99.7	554	14	US-10-136-227A-30	Sequence 32, Appl
24	3049.5	99.7	554	14	US-10-136-227A-32	Sequence 32, Appl
25	3049.5	99.7	554	14	US-10-112-881-30	Sequence 30, Appl
26	3049.5	99.7	554	14	US-10-112-881-32	Sequence 32, Appl
27	3049.5	99.7	554	14	US-10-112-881-32	Sequence 32, Appl
28	3047	99.6	559	9	US-09-981-649A-28	Sequence 28, Appl
29	3047	99.6	559	12	US-10-399-123-28	Sequence 28, Appl
30	3047	99.6	559	14	US-10-124-986-28	Sequence 28, Appl
31	3047	99.6	559	14	US-10-136-227A-28	Sequence 28, Appl
32	3047	99.6	559	14	US-10-112-881-28	Sequence 28, Appl
33	2769	90.5	537	9	US-09-981-649A-4	Sequence 4, Appli
34	2769	90.5	537	12	US-10-399-123-4	Sequence 4, Appli
35	2769	90.5	537	14	US-10-124-986-4	Sequence 4, Appli
36	2769	90.5	537	14	US-10-136-227A-4	Sequence 4, Appli
37	2769	90.5	537	14	US-10-112-881-4	Sequence 4, Appli
38	2757	90.1	502	9	US-09-981-649A-18	Sequence 18, Appl
39	2757	90.1	502	12	US-10-399-123-18	Sequence 18, Appl
40	2757	90.1	502	14	US-10-124-986-18	Sequence 18, Appl
41	2757	90.1	502	14	US-10-136-227A-18	Sequence 18, Appl
42	2757	90.1	502	14	US-10-112-881-18	Sequence 18, Appl
43	2428	79.3	551	10	US-09-866-050A-506	Sequence 506, App
44	2401.5	78.5	550	15	US-10-023-634-73	Sequence 73, Appl
45	1931	63.1	338	9	US-09-978-295A-119	Sequence 119, App

ALIGNMENTS

RESULT 1

US-09-981-649A-24
; Sequence 24, Application US/09981649A
; Patent No. US20020132250A1
; GENERAL INFORMATION:
; APPLICANT: Ford et al.
; TITLE OF INVENTION: EGF MOTIF PROTEIN, EGF6, MATERIALS AND METHODS
; FILE REFERENCE: 28110/37665
; CURRENT APPLICATION NUMBER: US/09/981,649A
; PRIOR FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: US 09/687,860
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 09/620,312
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: US 09/363,316
; PRIOR FILING DATE: 1999-07-28
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 24
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(553)
US-09-981-649A-24

Query Match 100.0%; Score 3060; DB 9; Length 553;
Best Local Similarity 100.0%; Pred. No. 3e-234;
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPLPWSLALPLLSSWAGGF...VDGVLLVSLGLASARQPCVCHYGTKLACCYGWRNSKGV 60
DB 1 MPLPWSLALPLLSSWAGGF...VDGVLLVSLGLASARQPCVCHYGTKLACCYGWRNSKGV 60
QY 61 CEATCEPGKFCGCVGNKRCFFPGYTGKTCSDVNECGMKPRPCOHRVCNTHGSKYKFC 120
DB 61 CEATCEPGKFCGCVGNKRCFFPGYTGKTCSDVNECGMKPRPCOHRVCNTHGSKYKFC 120

```
QY 121 LSGHMLPDCVNSRTCAMINCOYSCEDTEEGPQCLCPSSGLRLAPNGRDLIDECAS 180
Db 121 LSGHMLPDCVNSRTCAMINCOYSCEDTEEGPQCLCPSSGLRLAPNGRDLIDECAS 180
QY 181 GKVICPNRRCVNTFGSYCKCHIGFELQYISGRYDCIDINECTMDSHTCSHHANCFTQ 240
Db 181 GKVICPNRRCVNTFGSYCKCHIGFELQYISGRYDCIDINECTMDSHTCSHHANCFTQ 240
QY 241 GSPKCKCKQYKGNLRCSAIPENSVEVLRAFGTTIKDRIKLLAHKNSMKKAKIKNVT 300
Db 241 GSPKCKCKQYKGNLRCSAIPENSVEVLRAFGTTIKDRIKLLAHKNSMKKAKIKNVT 300
QY 301 PEPTRTPTPKVNLQPFNYEEIVSRGNSHGKGNNEKMEKLEDEKREKALKNDIEER 360
Db 301 PEPTRTPTPKVNLQPFNYEEIVSRGNSHGKGNNEKMEKLEDEKREKALKNDIEER 360
QY 361 SLRGDVFPPKVNAGFGLILVQRKALTSKLEHKDLNISVDCSFNHGICDWKQDREDDFD 420
Db 361 SLRGDVFPPKVNAGFGLILVQRKALTSKLEHKDLNISVDCSFNHGICDWKQDREDDFD 420
QY 421 WNPADRNAIGFYMAVPALAGHKDIDGLKLLLPDLQPSNFCLLFYRLAGDKVGLRV 480
Db 421 WNPADRNAIGFYMAVPALAGHKDIDGLKLLLPDLQPSNFCLLFYRLAGDKVGLRV 480
QY 481 FVKNSNNALAWKTTSEDEKWKTKIQLYQGTDTATKSIIFEAERGKGTGEIAVDGVLLV 540
Db 481 FVKNSNNALAWKTTSEDEKWKTKIQLYQGTDTATKSIIFEAERGKGTGEIAVDGVLLV 540
QY 541 SGLCPDLSLLSVDD 553
Db 541 SGLCPDLSLLSVDD 553

RESULT 2
US-09-796-753-14
; Sequence 14, Application US/09796753
; Publication No. US20030027998A1
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Sean A.
; TITLE OF INVENTION: SECRETED PROTEINS AND USES THEREOF
; FILE REFERENCE: 7853-227-999
; CURRENT APPLICATION NUMBER: US/09/796,753
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 09/183,175
; PRIOR FILING DATE: 1998-10-30
; PRIOR APPLICATION NUMBER: 09/223,094
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/223,546
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/224,246
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/259,388
; PRIOR FILING DATE: 1999-02-26
; PRIOR APPLICATION NUMBER: 60/122,458
; PRIOR FILING DATE: 1999-03-01
; PRIOR APPLICATION NUMBER: 09/312,359
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 09/336,536
; PRIOR FILING DATE: 1999-06-18
; PRIOR APPLICATION NUMBER: 09/342,687
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: 09/345,464
; PRIOR FILING DATE: 1999-06-30
; PRIOR APPLICATION NUMBER: 09/365,164
; PRIOR FILING DATE: 1999-07-30
; PRIOR APPLICATION NUMBER: 09/399,723
; PRIOR FILING DATE: 1999-09-20
; PRIOR APPLICATION NUMBER: 09/409,634
; PRIOR FILING DATE: 1999-09-30
; PRIOR APPLICATION NUMBER: 09/471,179
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: 09/474,071
```

```
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 09/474,072
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 09/514,010
; PRIOR FILING DATE: 2000-02-25
; PRIOR APPLICATION NUMBER: 09/516,745
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/572,002
; PRIOR FILING DATE: 2000-05-14
; PRIOR APPLICATION NUMBER: 09/597,993
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: 09/599,596
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 09/630,334
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: 09/606,565
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: 09/606,317
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: 09/665,666
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: 09/677,751
; PRIOR FILING DATE: 2000-09-30
; NUMBER OF SEQ ID NOS: 162
; SEQ ID NO 14
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-796-753-14
```

```
Query Match 100.0%; Score 3060; DB 10; Length 553;
Best Local Similarity 100.0%; Pred. No. 3e-234;
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPLPWSLALPLLLSWAGFGNAASARRHGLLASARQPGVCHYGTKLACCYGWRNRSGV 60
Db 1 MPLPWSLALPLLLSWAGFGNAASARRHGLLASARQPGVCHYGTKLACCYGWRNRSGV 60
QY 61 CEATCEPGCKFGECVGNKCRCPFGYTGKTCSDVNECGMKPRPCQHRVCNTHGSKYKFC 120
Db 61 CEATCEPGCKFGECVGNKCRCPFGYTGKTCSDVNECGMKPRPCQHRVCNTHGSKYKFC 120
QY 121 LSGHMLPDCVNSRTCAMINCOYSCEDTEEGPQCLCPSSGLRLAPNGRDLIDECAS 180
Db 121 LSGHMLPDCVNSRTCAMINCOYSCEDTEEGPQCLCPSSGLRLAPNGRDLIDECAS 180
QY 181 GKVICPNRRCVNTFGSYCKCHIGFELQYISGRYDCIDINECTMDSHTCSHHANCFTQ 240
Db 181 GKVICPNRRCVNTFGSYCKCHIGFELQYISGRYDCIDINECTMDSHTCSHHANCFTQ 240
QY 241 GSPKCKCKQYKGNLRCSAIPENSVEVLRAFGTTIKDRIKLLAHKNSMKKAKIKNVT 300
Db 241 GSPKCKCKQYKGNLRCSAIPENSVEVLRAFGTTIKDRIKLLAHKNSMKKAKIKNVT 300
QY 301 PEPTRTPTPKVNLQPFNYEEIVSRGNSHGKGNNEKMEKLEDEKREKALKNDIEER 360
Db 301 PEPTRTPTPKVNLQPFNYEEIVSRGNSHGKGNNEKMEKLEDEKREKALKNDIEER 360
QY 361 SLRGDVFPPKVNAGFGLILVQRKALTSKLEHKDLNISVDCSFNHGICDWKQDREDDFD 420
Db 361 SLRGDVFPPKVNAGFGLILVQRKALTSKLEHKDLNISVDCSFNHGICDWKQDREDDFD 420
QY 421 WNPADRNAIGFYMAVPALAGHKDIDGLKLLLPDLQPSNFCLLFYRLAGDKVGLRV 480
Db 421 WNPADRNAIGFYMAVPALAGHKDIDGLKLLLPDLQPSNFCLLFYRLAGDKVGLRV 480
QY 481 FVKNSNNALAWKTTSEDEKWKTKIQLYQGTDTATKSIIFEAERGKGTGEIAVDGVLLV 540
Db 481 FVKNSNNALAWKTTSEDEKWKTKIQLYQGTDTATKSIIFEAERGKGTGEIAVDGVLLV 540
QY 541 SGLCPDLSLLSVDD 553
Db 541 SGLCPDLSLLSVDD 553
```

```
RESULT 3
US-10-058-270A-102
; Sequence 102, Application US/10058270A
; Publication No. US20040029114A1
; GENERAL INFORMATION:
; APPLICANT: Mack, David H.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Afar, Daniel
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Breast Cancer, Compositions and
; Methods of Screening for Modulators of Breast Cancer
; FILE REFERENCE: 018501-005210US
; CURRENT APPLICATION NUMBER: US/10/058,270A
; CURRENT FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: US 60/263,965
; PRIOR FILING DATE: 2001-01-24
; PRIOR APPLICATION NUMBER: US 60/265,928
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: US 09/829,472
; PRIOR FILING DATE: 2001-04-09
; PRIOR APPLICATION NUMBER: US 60/282,698
; PRIOR FILING DATE: 2001-04-09
; PRIOR APPLICATION NUMBER: US 60/288,590
; PRIOR FILING DATE: 2001-05-04
; PRIOR APPLICATION NUMBER: US 60/294,443
; PRIOR FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 141
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 102
; TYPE: PRT
; LENGTH: 553
; ORGANISM: Homo sapiens
US-10-058-270A-102

Query Match      100.0%; Score 3060; DB 12; Length 553;
Best Local Similarity 100.0%; Pred. No. 3e-234;
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MFLPSLALPLLLSVAGFGNAASARHHGLLASARQPGVCHYGTKLACCYGWRNRSKGV 60
DB 1 MFLPSLALPLLLSVAGFGNAASARHHGLLASARQPGVCHYGTKLACCYGWRNRSKGV 60
QY 61 CEATCEPGCKFGCEVGNPKRCFPFGYGTCTSDVNECGMKPRPCQHRVCNTHGSKYKFC 120
DB 61 CEATCEPGCKFGCEVGNPKRCFPFGYGTCTSDVNECGMKPRPCQHRVCNTHGSKYKFC 120
QY 121 LSGHMLMPDATCVNSRTCAMINCYSCDETEBGPQCLCPSSGLRLAPNGRDCLDIDECAS 180
DB 121 LSGHMLMPDATCVNSRTCAMINCYSCDETEBGPQCLCPSSGLRLAPNGRDCLDIDECAS 180
QY 181 GKVICPNRRCVNTFGSYCKCHIGFELQYISGRYDCIDINECTMDSHTCSHHANCFNTQ 240
DB 181 GKVICPNRRCVNTFGSYCKCHIGFELQYISGRYDCIDINECTMDSHTCSHHANCFNTQ 240
QY 241 GSFKCKCKQGYKNGLRCSAIPENSVEVLRAPGTIKDRIKLLAHKNSMKKKAKIKNT 300
DB 241 GSFKCKCKQGYKNGLRCSAIPENSVEVLRAPGTIKDRIKLLAHKNSMKKKAKIKNT 300
QY 301 PEPTRTPTPKVNLQPNFYEEIVSRGNSHGKKGNEEKMEKEGLEDEKREKALKNIDIER 360
DB 301 PEPTRTPTPKVNLQPNFYEEIVSRGNSHGKKGNEEKMEKEGLEDEKREKALKNIDIER 360
QY 361 SLRGDVFPPKVNAGEFGLILVQRKALTSKLEHKDLNISVDCSFNHGICDMWKQDREDDFD 420
DB 361 SLRGDVFPPKVNAGEFGLILVQRKALTSKLEHKDLNISVDCSFNHGICDMWKQDREDDFD 420
QY 421 WNPADRDNAIGFYMAVPALAGHKDDIGRLKLLPDLQPOSNFCLLFDYRLAGDKVGLRV 480
DB 421 WNPADRDNAIGFYMAVPALAGHKDDIGRLKLLPDLQPOSNFCLLFDYRLAGDKVGLRV 480
QY 481 FVKNSNNALAWKTTSEDEKWKTKIQLYQGTDTATKSIIFEAERGKGTGEIADVGLLV 540
```

```
DB 481 FVKNSNNALAWKTTSEDEKWKTKIQLYQGTDTATKSIIFEAERGKGTGEIADVGLLV 540
QY 541 SGLCPDLSLLSVDD 553
DB 541 SGLCPDLSLLSVDD 553

RESULT 4
US-10-399-123-24
; Sequence 24, Application US/10399123
; Publication No. US20040059098A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq et al.
; TITLE OF INVENTION: EGF MOTIF PROTEIN, EGF6, MATERIALS AND METHODS
; FILE REFERENCE: 28110/37665
; CURRENT APPLICATION NUMBER: US/10/399,123
; CURRENT FILING DATE: 2003-04-14
; PRIOR APPLICATION NUMBER: US 09/687,860
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 09/620,312
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: US 09/363,316
; PRIOR FILING DATE: 1999-07-28
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 24
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(553)
US-10-399-123-24

Query Match      100.0%; Score 3060; DB 12; Length 553;
Best Local Similarity 100.0%; Pred. No. 3e-234;
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MFLPSLALPLLLSVAGFGNAASARHHGLLASARQPGVCHYGTKLACCYGWRNRSKGV 60
DB 1 MFLPSLALPLLLSVAGFGNAASARHHGLLASARQPGVCHYGTKLACCYGWRNRSKGV 60
QY 61 CEATCEPGCKFGCEVGNPKRCFPFGYGTCTSDVNECGMKPRPCQHRVCNTHGSKYKFC 120
DB 61 CEATCEPGCKFGCEVGNPKRCFPFGYGTCTSDVNECGMKPRPCQHRVCNTHGSKYKFC 120
QY 121 LSGHMLMPDATCVNSRTCAMINCYSCDETEBGPQCLCPSSGLRLAPNGRDCLDIDECAS 180
DB 121 LSGHMLMPDATCVNSRTCAMINCYSCDETEBGPQCLCPSSGLRLAPNGRDCLDIDECAS 180
QY 181 GKVICPNRRCVNTFGSYCKCHIGFELQYISGRYDCIDINECTMDSHTCSHHANCFNTQ 240
DB 181 GKVICPNRRCVNTFGSYCKCHIGFELQYISGRYDCIDINECTMDSHTCSHHANCFNTQ 240
QY 241 GSFKCKCKQGYKNGLRCSAIPENSVEVLRAPGTIKDRIKLLAHKNSMKKKAKIKNT 300
DB 241 GSFKCKCKQGYKNGLRCSAIPENSVEVLRAPGTIKDRIKLLAHKNSMKKKAKIKNT 300
QY 301 PEPTRTPTPKVNLQPNFYEEIVSRGNSHGKKGNEEKMEKEGLEDEKREKALKNIDIER 360
DB 301 PEPTRTPTPKVNLQPNFYEEIVSRGNSHGKKGNEEKMEKEGLEDEKREKALKNIDIER 360
QY 361 SLRGDVFPPKVNAGEFGLILVQRKALTSKLEHKDLNISVDCSFNHGICDMWKQDREDDFD 420
DB 361 SLRGDVFPPKVNAGEFGLILVQRKALTSKLEHKDLNISVDCSFNHGICDMWKQDREDDFD 420
QY 421 WNPADRDNAIGFYMAVPALAGHKDDIGRLKLLPDLQPOSNFCLLFDYRLAGDKVGLRV 480
DB 421 WNPADRDNAIGFYMAVPALAGHKDDIGRLKLLPDLQPOSNFCLLFDYRLAGDKVGLRV 480
QY 481 FVKNSNNALAWKTTSEDEKWKTKIQLYQGTDTATKSIIFEAERGKGTGEIADVGLLV 540
```

```
Db 481 FVKNSNNAWAKTTSEDEKWKTKIQLYQGTATKSIIFEAERGKGTGEIADVGVLLV 540
QY 541 SGLCPDSSLVSDD 553
Db 541 SGLCPDSSLVSDD 553

RESULT 5
US-10-124-986-24
; Sequence 24, Application US/10124986
; Publication No. US20030036508A1
; GENERAL INFORMATION:
; APPLICANT: Ford et al.
; TITLE OF INVENTION: EGF MOTIF PROTEIN, EGFL6, MATERIALS AND METHODS
; CURRENT APPLICATION NUMBER: US/10/124,986
; PRIOR FILING DATE: 2002-04-17
; PRIOR APPLICATION NUMBER: US 09/981,649
; PRIOR FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: US 09/687,860
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 09/620,312
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: US 09/363,316
; PRIOR FILING DATE: 1999-07-28
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 24
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)....(553)
US-10-124-986-24

Query Match 100.0%; Score 3060; DB 14; Length 553;
Best Local Similarity 100.0%; Pred. No. 3e-234;
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MFLPWSLALPLLSSWAGFGNNAASARHGLLASARQGVCHYGTKLACCYGRNRSKGV 60
Db 1 MFLPWSLALPLLSSWAGFGNNAASARHGLLASARQGVCHYGTKLACCYGRNRSKGV 60
QY 61 CEATCEPGCKFGEVGNPKRCFPFGYTGKTCSDQVNECGMKPRPCQHRVCVNTGHSYKFC 120
Db 61 CEATCEPGCKFGEVGNPKRCFPFGYTGKTCSDQVNECGMKPRPCQHRVCVNTGHSYKFC 120
QY 121 LSGHMLMPDATCVNSRTCAMINCOYSCEDTEEGPQCLCPSSGLRLAPNGRDCLDIDECAS 180
Db 121 LSGHMLMPDATCVNSRTCAMINCOYSCEDTEEGPQCLCPSSGLRLAPNGRDCLDIDECAS 180
QY 181 GKVICPNRRCVNTGSGYCKCHIGFELYISGRYDCIDINECTWDSHTCSHHANCFTQ 240
Db 181 GKVICPNRRCVNTGSGYCKCHIGFELYISGRYDCIDINECTWDSHTCSHHANCFTQ 240
QY 241 GSFCKCKQGYKGNGLRCSAIPENSVEVLRAPGTIKDRIKLLAHKNSMCKKAKIKNVT 300
Db 241 GSFCKCKQGYKGNGLRCSAIPENSVEVLRAPGTIKDRIKLLAHKNSMCKKAKIKNVT 300
QY 301 PEPTRTPTPKVNLQPNFYEEIVSRGNSHGKKGNEEKMEKGELEDEKREKALKNDIEER 360
Db 301 PEPTRTPTPKVNLQPNFYEEIVSRGNSHGKKGNEEKMEKGELEDEKREKALKNDIEER 360
QY 361 SLRGDVFPKVNAGEFGLILVQRKALTSKLEHKDLNISVDCSFNHHGICDWKQDREDDFD 420
Db 361 SLRGDVFPKVNAGEFGLILVQRKALTSKLEHKDLNISVDCSFNHHGICDWKQDREDDFD 420
QY 421 WNPADRDNAIGFYMAVPALAGHKDIGRLKLLLPDLPQSNFCLLDYRLAGDKVKGKLRV 480
Db 421 WNPADRDNAIGFYMAVPALAGHKDIGRLKLLLPDLPQSNFCLLDYRLAGDKVKGKLRV 480
QY 481 FVKNSNNAWAKTTSEDEKWKTKIQLYQGTATKSIIFEAERGKGTGEIADVGVLLV 540
```

```
Db 481 FVKNSNNAWAKTTSEDEKWKTKIQLYQGTATKSIIFEAERGKGTGEIADVGVLLV 540
QY 541 SGLCPDSSLVSDD 553
Db 541 SGLCPDSSLVSDD 553

RESULT 6
US-10-136-227A-24
; Sequence 24, Application US/10136227A
; Publication No. US20030165886A1
; GENERAL INFORMATION:
; APPLICANT: Ford et al.
; TITLE OF INVENTION: EGF MOTIF PROTEIN, EGFL6, MATERIALS AND METHODS
; CURRENT APPLICATION NUMBER: US/10/136,227A
; PRIOR FILING DATE: 2002-05-01
; PRIOR APPLICATION NUMBER: US 09/981,649
; PRIOR FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: US 09/687,860
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 09/620,312
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: US 09/363,316
; PRIOR FILING DATE: 1999-07-28
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 24
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)....(553)
US-10-136-227A-24

Query Match 100.0%; Score 3060; DB 14; Length 553;
Best Local Similarity 100.0%; Pred. No. 3e-234;
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MFLPWSLALPLLSSWAGFGNNAASARHGLLASARQGVCHYGTKLACCYGRNRSKGV 60
Db 1 MFLPWSLALPLLSSWAGFGNNAASARHGLLASARQGVCHYGTKLACCYGRNRSKGV 60
QY 61 CEATCEPGCKFGEVGNPKRCFPFGYTGKTCSDQVNECGMKPRPCQHRVCVNTGHSYKFC 120
Db 61 CEATCEPGCKFGEVGNPKRCFPFGYTGKTCSDQVNECGMKPRPCQHRVCVNTGHSYKFC 120
QY 121 LSGHMLMPDATCVNSRTCAMINCOYSCEDTEEGPQCLCPSSGLRLAPNGRDCLDIDECAS 180
Db 121 LSGHMLMPDATCVNSRTCAMINCOYSCEDTEEGPQCLCPSSGLRLAPNGRDCLDIDECAS 180
QY 181 GKVICPNRRCVNTGSGYCKCHIGFELYISGRYDCIDINECTWDSHTCSHHANCFTQ 240
Db 181 GKVICPNRRCVNTGSGYCKCHIGFELYISGRYDCIDINECTWDSHTCSHHANCFTQ 240
QY 241 GSFCKCKQGYKGNGLRCSAIPENSVEVLRAPGTIKDRIKLLAHKNSMCKKAKIKNVT 300
Db 241 GSFCKCKQGYKGNGLRCSAIPENSVEVLRAPGTIKDRIKLLAHKNSMCKKAKIKNVT 300
QY 301 PEPTRTPTPKVNLQPNFYEEIVSRGNSHGKKGNEEKMEKGELEDEKREKALKNDIEER 360
Db 301 PEPTRTPTPKVNLQPNFYEEIVSRGNSHGKKGNEEKMEKGELEDEKREKALKNDIEER 360
QY 361 SLRGDVFPKVNAGEFGLILVQRKALTSKLEHKDLNISVDCSFNHHGICDWKQDREDDFD 420
Db 361 SLRGDVFPKVNAGEFGLILVQRKALTSKLEHKDLNISVDCSFNHHGICDWKQDREDDFD 420
QY 421 WNPADRDNAIGFYMAVPALAGHKDIGRLKLLLPDLPQSNFCLLDYRLAGDKVKGKLRV 480
Db 421 WNPADRDNAIGFYMAVPALAGHKDIGRLKLLLPDLPQSNFCLLDYRLAGDKVKGKLRV 480
```

QY 481 FVKNSSNALAWKTTSEDEKWKTKIQLYQGTDTATKSIIFEAERGKGTGEIAVGVLLV 540
 Db |||||
 QY 481 FVKNSSNALAWKTTSEDEKWKTKIQLYQGTDTATKSIIFEAERGKGTGEIAVGVLLV 540
 Db |||||
 QY 541 SGLCPDLSLLSVDD 553
 Db |||||
 QY 541 SGLCPDLSLLSVDD 553
 Db |||||
 RESULT 7
 US-10-112-881-24
 ; Sequence 24, Application US/10112881
 ; Publication No. US20030166909A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Ford et al.
 ; TITLE OF INVENTION: EGF MOTIF PROTEIN, EGFL6, MATERIALS AND METHODS
 ; FILE REFERENCE: 28110/38363
 ; CURRENT APPLICATION NUMBER: US/10/112,881
 ; CURRENT FILING DATE: 2002-03-29
 ; PRIOR APPLICATION NUMBER: US 09/981,649
 ; PRIOR FILING DATE: 2001-10-15
 ; PRIOR APPLICATION NUMBER: US 09/687,860
 ; PRIOR FILING DATE: 2000-10-13
 ; PRIOR APPLICATION NUMBER: US 09/620,312
 ; PRIOR FILING DATE: 2000-07-19
 ; PRIOR APPLICATION NUMBER: US 09/363,316
 ; PRIOR FILING DATE: 1999-07-28
 ; PRIOR APPLICATION NUMBER: US 09/249,697
 ; PRIOR FILING DATE: 1999-02-12
 ; PRIOR APPLICATION NUMBER: US 08/968,800
 ; PRIOR FILING DATE: 1997-11-22
 ; NUMBER OF SEQ ID NOS: 32
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 24
 ; LENGTH: 553
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: VARIANT
 ; LOCATION: (1)...(553)
 US-10-112-881-24
 Query Match 100.0%; Score 3060; DB 14; Length 553;
 Best Local Similarity 100.0%; Pred. No. 3e-234;
 Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MFLPWSLALPLLISWVAGFGNNAASARHGLLASARQPGVCHYGTKLACCYGWRNRSKGV 60
 Db |||||
 QY 61 CEATCEPGCKFGECVGNPKRCFPFGYTGKTCSDQVNECGMKPRPCQHRVNTHGSYKFC 120
 Db |||||
 QY 121 LSGHMLPDTATCNSTRCAINCOYSCDETEEGPQCLCPSSGLRLAPNGRDCDIDIBECAS 180
 Db |||||
 QY 181 GKVICPNRRCVNTFGSYCKCHIGFELQVTSYGRYDCIDINECTMDSHTCSSHANCFNTQ 240
 Db |||||
 QY 241 GSPFKCKQYKGNGLRCSAIPENSVEKVLAPGTIKDRIKKLLAHKNSMKKAKIKNVT 300
 Db |||||
 QY 301 PEPTRTTPKVNLOPFNYEETVSRGSHGCKGKNEEKMEGLEDEKREKALKNDIEER 360
 Db |||||
 QY 361 SLRGDVFPPKNEAEGFLILVQRKALTSLKLEHKLNLISVDCSPFNHGI CDWKQDREDDFD 420
 Db |||||

QY 421 WNPADRDNAIGFYMAVPALAGHKDIGRLKLLPDLQPOSNFCLLFDYRLAGDKVGLRV 480
 Db |||||
 QY 481 FVKNSSNALAWKTTSEDEKWKTKIQLYQGTDTATKSIIFEAERGKGTGEIAVGVLLV 540
 Db |||||
 QY 541 SGLCPDLSLLSVDD 553
 Db |||||
 QY 541 SGLCPDLSLLSVDD 553
 Db |||||
 RESULT 8
 US-10-295-027-494
 ; Sequence 494, Application US/10295027
 ; Publication No. US20030232350A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Afar, Daniel
 ; APPLICANT: Aziz, Natasha
 ; APPLICANT: Ginsberg, Wendy M.
 ; APPLICANT: Gish, Kurt C.
 ; APPLICANT: Glynn, Richard
 ; APPLICANT: Hevezi, Peter A.
 ; APPLICANT: Mack, David H.
 ; APPLICANT: Murray, Richard
 ; APPLICANT: Watson, Susan R.
 ; APPLICANT: Eos Biotechnology, Inc.
 ; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
 ; TITLE OF INVENTION: Methods of Screening for Modulators of Cancer
 ; FILE REFERENCE: 018501-01250005
 ; CURRENT APPLICATION NUMBER: US/10/295,027
 ; CURRENT FILING DATE: 2002-11-13
 ; PRIOR APPLICATION NUMBER: US 09/663,733
 ; PRIOR FILING DATE: 2000-09-15
 ; PRIOR APPLICATION NUMBER: US 60/350,666
 ; PRIOR FILING DATE: 2001-11-13
 ; PRIOR APPLICATION NUMBER: US 60/335,394
 ; PRIOR FILING DATE: 2001-11-15
 ; PRIOR APPLICATION NUMBER: US 60/332,464
 ; PRIOR FILING DATE: 2001-11-21
 ; PRIOR APPLICATION NUMBER: US 60/334,393
 ; PRIOR FILING DATE: 2001-11-29
 ; PRIOR APPLICATION NUMBER: US 60/340,376
 ; PRIOR FILING DATE: 2001-12-14
 ; PRIOR APPLICATION NUMBER: US 60/347,211
 ; PRIOR FILING DATE: 2002-01-08
 ; PRIOR APPLICATION NUMBER: US 60/347,349
 ; PRIOR FILING DATE: 2002-01-10
 ; PRIOR APPLICATION NUMBER: US 60/355,250
 ; PRIOR FILING DATE: 2002-02-08
 ; PRIOR APPLICATION NUMBER: US 60/356,714
 ; PRIOR FILING DATE: 2002-02-13
 ; Remaining Prior Application data removed - See File Wrapper or PALM.
 ; NUMBER OF SEQ ID NOS: 1386
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 494
 ; LENGTH: 553
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-295-027-494
 Query Match 100.0%; Score 3060; DB 15; Length 553;
 Best Local Similarity 100.0%; Pred. No. 3e-234;
 Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MFLPWSLALPLLISWVAGFGNNAASARHGLLASARQPGVCHYGTKLACCYGWRNRSKGV 60
 Db |||||
 QY 61 CEATCEPGCKFGECVGNPKRCFPFGYTGKTCSDQVNECGMKPRPCQHRVNTHGSYKFC 120
 Db |||||

```
Db 61 CEATCEPGCKGECVGNKCRCPGYTGKTCSDVNECGMKPRPCQHRVCVTHGSYKFC 120
Qy 121 LSGHMLPDCVNSRTCAMINQYSCBDETEGPOCLCPSSGLRLAPNGRCLDIDECAS 180
Db 121 LSGHMLPDCVNSRTCAMINQYSCBDETEGPOCLCPSSGLRLAPNGRCLDIDECAS 180
Qy 181 GKVICPYNRRVCNTFGSYCKCHIGFELQYISGRYDCIDINECTMDSHTCSHHANCFTQ 240
Db 181 GKVICPYNRRVCNTFGSYCKCHIGFELQYISGRYDCIDINECTMDSHTCSHHANCFTQ 240
Qy 241 GSFKCKCKQYKGNGLRCSAIPENSVEVLRAPIGTDRIKLLAHKNSMKKAKIKNVT 300
Db 241 GSFKCKCKQYKGNGLRCSAIPENSVEVLRAPIGTDRIKLLAHKNSMKKAKIKNVT 300
Qy 301 PEPTPTPKVNLQPNFYEEIVSRGNSHGKKGNEERKKGLEDEKREKALKNDIEER 360
Db 301 PEPTPTPKVNLQPNFYEEIVSRGNSHGKKGNEERKKGLEDEKREKALKNDIEER 360
Qy 361 SLRGDVFFPKVNEAGEFGLILVQKALTSKLEHKDLNLSVDCSFNNGICDMKQDREDDFD 420
Db 361 SLRGDVFFPKVNEAGEFGLILVQKALTSKLEHKDLNLSVDCSFNNGICDMKQDREDDFD 420
Qy 421 WNPADRDNAIGFYMAVPALAGHKDIGRLKLLPDLQPSNFCLLFDYRLAGDKVGLRV 480
Db 421 WNPADRDNAIGFYMAVPALAGHKDIGRLKLLPDLQPSNFCLLFDYRLAGDKVGLRV 480
Qy 481 FVKNSNNALAWKETSDEKWKTKIQLYQGTDTATKSIIFEABRGKGTGEIADVGVLLV 540
Db 481 FVKNSNNALAWKETSDEKWKTKIQLYQGTDTATKSIIFEABRGKGTGEIADVGVLLV 540
Qy 541 SGLCPDLSLLSVD 553
Db 541 SGLCPDLSLLSVD 553
```

RESULT 9

```
US-10-295-027-812
; Sequence 812, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynn, Richard
; APPLICANT: Hevezi, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
; TITLE OF INVENTION: Methods of Screening for Modulators of Cancer
; FILE REFERENCE: 018501-01250005
; CURRENT FILING DATE: 2002-11-13
; PRIOR FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR FILING DATE: 2002-02-08
```

```
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 812
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-295-027-812

Query Match 100.0%; Score 3060; DB 15; Length 553;
Best Local Similarity 100.0%; Pred. No. 3e-234;
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MPLPWSLALPLLISWAGFGNAASARHHGLLASARQPGVCHYGTKLACCYGWRNRSKV 60
Db 1 MPLPWSLALPLLISWAGFGNAASARHHGLLASARQPGVCHYGTKLACCYGWRNRSKV 60
Qy 61 CEATCEPGCKGECVGNKCRCPGYTGKTCSDVNECGMKPRPCQHRVCVTHGSYKFC 120
Db 61 CEATCEPGCKGECVGNKCRCPGYTGKTCSDVNECGMKPRPCQHRVCVTHGSYKFC 120
Qy 121 LSGHMLPDCVNSRTCAMINQYSCBDETEGPOCLCPSSGLRLAPNGRCLDIDECAS 180
Db 121 LSGHMLPDCVNSRTCAMINQYSCBDETEGPOCLCPSSGLRLAPNGRCLDIDECAS 180
Qy 181 GKVICPYNRRVCNTFGSYCKCHIGFELQYISGRYDCIDINECTMDSHTCSHHANCFTQ 240
Db 181 GKVICPYNRRVCNTFGSYCKCHIGFELQYISGRYDCIDINECTMDSHTCSHHANCFTQ 240
Qy 241 GSFKCKCKQYKGNGLRCSAIPENSVEVLRAPIGTDRIKLLAHKNSMKKAKIKNVT 300
Db 241 GSFKCKCKQYKGNGLRCSAIPENSVEVLRAPIGTDRIKLLAHKNSMKKAKIKNVT 300
Qy 301 PEPTPTPKVNLQPNFYEEIVSRGNSHGKKGNEERKKGLEDEKREKALKNDIEER 360
Db 301 PEPTPTPKVNLQPNFYEEIVSRGNSHGKKGNEERKKGLEDEKREKALKNDIEER 360
Qy 361 SLRGDVFFPKVNEAGEFGLILVQKALTSKLEHKDLNLSVDCSFNNGICDMKQDREDDFD 420
Db 361 SLRGDVFFPKVNEAGEFGLILVQKALTSKLEHKDLNLSVDCSFNNGICDMKQDREDDFD 420
Qy 421 WNPADRDNAIGFYMAVPALAGHKDIGRLKLLPDLQPSNFCLLFDYRLAGDKVGLRV 480
Db 421 WNPADRDNAIGFYMAVPALAGHKDIGRLKLLPDLQPSNFCLLFDYRLAGDKVGLRV 480
Qy 481 FVKNSNNALAWKETSDEKWKTKIQLYQGTDTATKSIIFEABRGKGTGEIADVGVLLV 540
Db 481 FVKNSNNALAWKETSDEKWKTKIQLYQGTDTATKSIIFEABRGKGTGEIADVGVLLV 540
Qy 541 SGLCPDLSLLSVD 553
Db 541 SGLCPDLSLLSVD 553
```

RESULT 10

```
US-10-295-027-841
; Sequence 841, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynn, Richard
; APPLICANT: Hevezi, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
; TITLE OF INVENTION: Methods of Screening for Modulators of Cancer
```

```
/ FILE REFERENCE: 018501-012500US
/ CURRENT APPLICATION NUMBER: US/10/295,027
/ CURRENT FILING DATE: 2002-11-13
/ PRIOR APPLICATION NUMBER: US 09/663,733
/ PRIOR FILING DATE: 2000-09-15
/ PRIOR APPLICATION NUMBER: US 60/350,666
/ PRIOR FILING DATE: 2001-11-13
/ PRIOR APPLICATION NUMBER: US 60/335,394
/ PRIOR FILING DATE: 2001-11-15
/ PRIOR APPLICATION NUMBER: US 60/332,464
/ PRIOR FILING DATE: 2001-11-21
/ PRIOR APPLICATION NUMBER: US 60/334,393
/ PRIOR FILING DATE: 2001-11-29
/ PRIOR APPLICATION NUMBER: US 60/340,376
/ PRIOR FILING DATE: 2001-12-14
/ PRIOR APPLICATION NUMBER: US 60/347,211
/ PRIOR FILING DATE: 2002-01-08
/ PRIOR APPLICATION NUMBER: US 60/347,349
/ PRIOR FILING DATE: 2002-01-10
/ PRIOR APPLICATION NUMBER: US 60/355,250
/ PRIOR FILING DATE: 2002-02-08
/ PRIOR APPLICATION NUMBER: US 60/356,714
/ PRIOR FILING DATE: 2002-02-13
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 1386
/ SOFTWARE: Patentin Ver. 2.1
/ SEQ ID NO 841
/ LENGTH: 553
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-295-027-841

Query Match 100.0%; Score 3060; DB 15; Length 553;
Best Local Similarity 100.0%; Pred. No. 3e-234;
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MFLPNSLALPLLWSVAGFGNNAASARHHGLLASARQGVCHYGTKLACCYGWRNRSKV 60
DB 1 MFLPNSLALPLLWSVAGFGNNAASARHHGLLASARQGVCHYGTKLACCYGWRNRSKV 60
QY 61 CEATCEPGCKFGCEGVGNKRCFCPGYTGKTCSDQVNECGMKPRPCQHRCVNTHGSKFC 120
DB 61 CEATCEPGCKFGCEGVGNKRCFCPGYTGKTCSDQVNECGMKPRPCQHRCVNTHGSKFC 120
QY 121 LSGHMLMPDATCVNSRTCAMINCOYSCDETEGPGQCLCPSSGLRLAPNGRDCLDIDECAS 180
DB 121 LSGHMLMPDATCVNSRTCAMINCOYSCDETEGPGQCLCPSSGLRLAPNGRDCLDIDECAS 180
QY 181 GKVICPNRRCVNTFGSYCKCHIGFELQYISGRYDCIDINECTMDSHTCSHHANCFNTQ 240
DB 181 GKVICPNRRCVNTFGSYCKCHIGFELQYISGRYDCIDINECTMDSHTCSHHANCFNTQ 240
QY 241 GSFCKCKQGYKNGLRCSAIPENSVEVLRAPGTIKDRIKLLAHKNSMKKAKIKNVT 300
DB 241 GSFCKCKQGYKNGLRCSAIPENSVEVLRAPGTIKDRIKLLAHKNSMKKAKIKNVT 300
QY 301 PEPTRTPTKVNLPFNVEEIVSRGNSHGKKGNEEKMEKGELEDEKREKALKNDIEER 360
DB 301 PEPTRTPTKVNLPFNVEEIVSRGNSHGKKGNEEKMEKGELEDEKREKALKNDIEER 360
QY 361 SURGDVFFPKVNEAGEFGLILVORKALTSKLEHKDNLNISVDCSFNNGICDWKQDREDDFD 420
DB 361 SURGDVFFPKVNEAGEFGLILVORKALTSKLEHKDNLNISVDCSFNNGICDWKQDREDDFD 420
QY 421 WNPADRDNAIGFYMAVPALAGHKDDIGRLKLLPDLQPOSNFCLLFDYRLAGDKVGLRV 480
DB 421 WNPADRDNAIGFYMAVPALAGHKDDIGRLKLLPDLQPOSNFCLLFDYRLAGDKVGLRV 480
QY 481 FVKNSNNALAWKTTSEDEKWKTKIQLYQGTATKSIIFEARGKGTGETAIVDGLLV 540
DB 481 FVKNSNNALAWKTTSEDEKWKTKIQLYQGTATKSIIFEARGKGTGETAIVDGLLV 540
QY 541 SGLCPDLSLLSVDD 553

/ FILE REFERENCE: 018501-002420US
/ CURRENT APPLICATION NUMBER: US/10/173,999
/ Publication No. US20040005563A1
/ GENERAL INFORMATION:
/ APPLICANT: Mack, David H.
/ APPLICANT: Gish, Kurt C.
/ APPLICANT: Eos Biotechnology, Inc.
/ TITLE OF INVENTION: Methods of Diagnosis of Ovarian Cancer, Compositions
/ TITLE OF INVENTION: and Methods of Screening for Modulators of Ovarian
/ TITLE OF INVENTION: Cancer
/ FILE REFERENCE: 018501-002420US
/ CURRENT APPLICATION NUMBER: US/10/173,999
/ CURRENT FILING DATE: 2002-06-17
/ PRIOR APPLICATION NUMBER: US 60/299,234
/ PRIOR FILING DATE: 2001-06-18
/ PRIOR APPLICATION NUMBER: US 60/315,287
/ PRIOR FILING DATE: 2001-08-27
/ PRIOR APPLICATION NUMBER: US 60/350,666
/ PRIOR FILING DATE: 2001-11-13
/ PRIOR APPLICATION NUMBER: US 60/372,246
/ PRIOR FILING DATE: 2001-04-12
/ NUMBER OF SEQ ID NOS: 163
/ SOFTWARE: Patentin Ver. 2.1
/ SEQ ID NO 46
/ LENGTH: 553
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-173-999-46

Query Match 100.0%; Score 3060; DB 15; Length 553;
Best Local Similarity 100.0%; Pred. No. 3e-234;
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MFLPNSLALPLLWSVAGFGNNAASARHHGLLASARQGVCHYGTKLACCYGWRNRSKV 60
DB 1 MFLPNSLALPLLWSVAGFGNNAASARHHGLLASARQGVCHYGTKLACCYGWRNRSKV 60
QY 61 CEATCEPGCKFGCEGVGNKRCFCPGYTGKTCSDQVNECGMKPRPCQHRCVNTHGSKFC 120
DB 61 CEATCEPGCKFGCEGVGNKRCFCPGYTGKTCSDQVNECGMKPRPCQHRCVNTHGSKFC 120
QY 121 LSGHMLMPDATCVNSRTCAMINCOYSCDETEGPGQCLCPSSGLRLAPNGRDCLDIDECAS 180
DB 121 LSGHMLMPDATCVNSRTCAMINCOYSCDETEGPGQCLCPSSGLRLAPNGRDCLDIDECAS 180
QY 181 GKVICPNRRCVNTFGSYCKCHIGFELQYISGRYDCIDINECTMDSHTCSHHANCFNTQ 240
DB 181 GKVICPNRRCVNTFGSYCKCHIGFELQYISGRYDCIDINECTMDSHTCSHHANCFNTQ 240
QY 241 GSFCKCKQGYKNGLRCSAIPENSVEVLRAPGTIKDRIKLLAHKNSMKKAKIKNVT 300
DB 241 GSFCKCKQGYKNGLRCSAIPENSVEVLRAPGTIKDRIKLLAHKNSMKKAKIKNVT 300
QY 301 PEPTRTPTKVNLPFNVEEIVSRGNSHGKKGNEEKMEKGELEDEKREKALKNDIEER 360
DB 301 PEPTRTPTKVNLPFNVEEIVSRGNSHGKKGNEEKMEKGELEDEKREKALKNDIEER 360
QY 361 SURGDVFFPKVNEAGEFGLILVORKALTSKLEHKDNLNISVDCSFNNGICDWKQDREDDFD 420
DB 361 SURGDVFFPKVNEAGEFGLILVORKALTSKLEHKDNLNISVDCSFNNGICDWKQDREDDFD 420
QY 421 WNPADRDNAIGFYMAVPALAGHKDDIGRLKLLPDLQPOSNFCLLFDYRLAGDKVGLRV 480
DB 421 WNPADRDNAIGFYMAVPALAGHKDDIGRLKLLPDLQPOSNFCLLFDYRLAGDKVGLRV 480
QY 481 FVKNSNNALAWKTTSEDEKWKTKIQLYQGTATKSIIFEARGKGTGETAIVDGLLV 540
DB 481 FVKNSNNALAWKTTSEDEKWKTKIQLYQGTATKSIIFEARGKGTGETAIVDGLLV 540
QY 541 SGLCPDLSLLSVDD 553
```


QY 541 SGLCPDLSLSD 553
Db 541 SGLCPDLSLSD 553

RESULT 12
US-10-188-832-189
; Sequence 189, Application US/10188832
; Publication No. US20040076955A1
; GENERAL INFORMATION:
; APPLICANT: Mack, David H.
; APPLICANT: Aziz, Natasha
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Bladder Cancer, Compositions
; TITLE OF INVENTION: and Methods of Screening for Modulators of Bladder
; TITLE OF INVENTION: Cancer
; FILE REFERENCE: 018501-0023300S
; CURRENT APPLICATION NUMBER: US/10/188,832
; CURRENT FILING DATE: 2002-11-22
; PRIOR APPLICATION NUMBER: US 60/302,814
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: US 60/310,099
; PRIOR FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US 60/343,705
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/372,246
; PRIOR FILING DATE: 2002-04-12
; NUMBER OF SEQ ID NOS: 207
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 189
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-188-832-189

Query Match 100.0%; Score 3060; DB 16; Length 553;
Best Local Similarity 100.0%; Pred. No. 3e-234;
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MFLPWSLALPLLSSVAGGFGNAASARHGLLASARQPGVCHYGTKLACCYGWRNRSKV 60
Db 1 MFLPWSLALPLLSSVAGGFGNAASARHGLLASARQPGVCHYGTKLACCYGWRNRSKV 60
QY 61 CEATCEPCKGECVGNKRCFCPGYTGKTCSDVNECGMKPRPCQHRVNTHGSYKFC 120
Db 61 CEATCEPCKGECVGNKRCFCPGYTGKTCSDVNECGMKPRPCQHRVNTHGSYKFC 120
QY 121 LSGHMLMPDATCVNSRTCAMINCOYSCEDTEBGPQCLCPSSGLRLAPNGRDCLDIDECAS 180
Db 121 LSGHMLMPDATCVNSRTCAMINCOYSCEDTEBGPQCLCPSSGLRLAPNGRDCLDIDECAS 180
QY 181 GKVICPNRRCVNTFGSYCKCHIGFELQYISGRYDCIDINECTMDSHTCSHHANCFTQ 240
Db 181 GKVICPNRRCVNTFGSYCKCHIGFELQYISGRYDCIDINECTMDSHTCSHHANCFTQ 240
QY 241 GSFCKCKQGYKGNLRCSSAIPENSVEKVLAPGTIKDRIKLLAHKNSMKKAKIKNVT 300
Db 241 GSFCKCKQGYKGNLRCSSAIPENSVEKVLAPGTIKDRIKLLAHKNSMKKAKIKNVT 300
QY 301 PEPTRTPTKVNLPQFNYYEIVSRGNSHGKKGNEEKMEGLEDEKREKALKNDIEER 360
Db 301 PEPTRTPTKVNLPQFNYYEIVSRGNSHGKKGNEEKMEGLEDEKREKALKNDIEER 360
QY 361 SURGDVFPKVNAGEFGLILVQRKALTSKLEHKDNLISVDCSFNHIICDWKQDREDDFD 420
Db 361 SURGDVFPKVNAGEFGLILVQRKALTSKLEHKDNLISVDCSFNHIICDWKQDREDDFD 420
QY 421 WNPADRDNAIGFYMAVPALAGHKDIDRLKLLPDLQPSNFCLLDYRLAGDKVGLRV 480
Db 421 WNPADRDNAIGFYMAVPALAGHKDIDRLKLLPDLQPSNFCLLDYRLAGDKVGLRV 480

QY 481 FVKNNSNALAWEXTTSEDEKWKTKIQLYQGTDATKSIIFEARGRGKTGEIAVDGVLV 540
Db 481 FVKNNSNALAWEXTTSEDEKWKTKIQLYQGTDATKSIIFEARGRGKTGEIAVDGVLV 540
QY 541 SGLCPDLSLSD 553
Db 541 SGLCPDLSLSD 553
RESULT 13
US-09-981-649A-6
; Sequence 6, Application US/09981649A
; Patent No. US20020132250A1
; GENERAL INFORMATION:
; APPLICANT: Ford et al.
; TITLE OF INVENTION: EGF MOTIF PROTEIN, EGFL6, MATERIALS AND METHODS
; FILE REFERENCE: 28110/37665
; CURRENT APPLICATION NUMBER: US/09/981,649A
; CURRENT FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: US 09/687,860
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 09/620,312
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: US 09/363,316
; PRIOR FILING DATE: 1999-07-28
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (357)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-981-649A-6

Query Match 99.7%; Score 3050; DB 9; Length 553;
Best Local Similarity 99.6%; Pred. No. 1.9e-233;
Matches 551; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 MFLPWSLALPLLSSVAGGFGNAASARHGLLASARQPGVCHYGTKLACCYGWRNRSKV 60
Db 1 MFLPWSLALPLLSSVAGGFGNAASARHGLLASARQPGVCHYGTKLACCYGWRNRSKV 60
QY 61 CEATCEPCKGECVGNKRCFCPGYTGKTCSDVNECGMKPRPCQHRVNTHGSYKFC 120
Db 61 CEATCEPCKGECVGNKRCFCPGYTGKTCSDVNECGMKPRPCQHRVNTHGSYKFC 120
QY 121 LSGHMLMPDATCVNSRTCAMINCOYSCEDTEBGPQCLCPSSGLRLAPNGRDCLDIDECAS 180
Db 121 LSGHMLMPDATCVNSRTCAMINCOYSCEDTEBGPQCLCPSSGLRLAPNGRDCLDIDECAS 180
QY 181 GKVICPNRRCVNTFGSYCKCHIGFELQYISGRYDCIDINECTMDSHTCSHHANCFTQ 240
Db 181 GKVICPNRRCVNTFGSYCKCHIGFELQYISGRYDCIDINECTMDSHTCSHHANCFTQ 240
QY 241 GSFCKCKQGYKGNLRCSSAIPENSVEKVLAPGTIKDRIKLLAHKNSMKKAKIKNVT 300
Db 241 GSFCKCKQGYKGNLRCSSAIPENSVEKVLAPGTIKDRIKLLAHKNSMKKAKIKNVT 300
QY 301 PEPTRTPTKVNLPQFNYYEIVSRGNSHGKKGNEEKMEGLEDEKREKALKNDIEER 360
Db 301 PEPTRTPTKVNLPQFNYYEIVSRGNSHGKKGNEEKMEGLEDEKREKALKNDIEER 360
QY 361 SURGDVFPKVNAGEFGLILVQRKALTSKLEHKDNLISVDCSFNHIICDWKQDREDDFD 420
Db 361 SURGDVFPKVNAGEFGLILVQRKALTSKLEHKDNLISVDCSFNHIICDWKQDREDDFD 420
QY 421 WNPADRDNAIGFYMAVPALAGHKDIDRLKLLPDLQPSNFCLLDYRLAGDKVGLRV 480
Db 421 WNPADRDNAIGFYMAVPALAGHKDIDRLKLLPDLQPSNFCLLDYRLAGDKVGLRV 480

QY	481	FVYNNNALAWKTTSEDEKKWTKIQLYQGTDTATKSII	FEAERGGKGTGETIADGVLLV	540
DB	481	FVYNNNALAWKTTSEDEKKWTKIQLYQGTDTATKSII	FEAERGGKGTGETIADGVLLV	540
QY	541	SGLCPSDLSLLSVDD	553	
DB	541	SGLCPSDLSLLSVDD	553	

```

RESULT 14
US-10-399-123-6
; Sequence 6, Application US/10399123
; Publication NO. US20040059098A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq et al.
; TITLE OF INVENTION: EGF MOTIF PROTEIN, EGFL6, MATERIALS AND METHODS
; FILE REFERENCE: 28110/37865
; CURRENT APPLICATION NUMBER: US/10/399,123
; CURRENT FILING DATE: 2003-04-14
; PRIOR APPLICATION NUMBER: US 09/687,860
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 09/620,312
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: US 09/363,316
; PRIOR FILING DATE: 1999-07-28
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows version 3.0
; SEQ ID NO 6
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(553)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-10-399-123-6

```

Qy	481	FVKSNNALAWKETTSEDEKWKTKIQLYQGTDAKSIIFPAERGKGTGSIADVGVLLV	540
Db	481	FVKSNNALAWKETTSEDEKWKTKIQLYQGTDAKSIIFPAERGKGTGSIADVGVLLV	540
Qy	541	SGLCPSDILLSVDD	553
Db	541	SGLCPSDILLSVDD	553

```

RESULT 15
US-10-124-986-6
; Sequence 6, Application US/10124986
; Publication NO. US20030036508A1
; GENERAL INFORMATION:
; APPLICANT: Ford et al.
; TITLE OF INVENTION: EGF MOTIF PROTEIN, EGFL6, MATERIALS AND METHODS
; FILE REFERENCE: 28110/37958
; CURRENT APPLICATION NUMBER: US/10/124,986
; CURRENT FILING DATE: 2002-04-17
; PRIOR APPLICATION NUMBER: US 09/981,649
; PRIOR FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: US 09/687,860
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 09/620,312
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: US 09/363,316
; PRIOR FILING DATE: 1999-07-28
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (357)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-10-124-986-6

```

Db	421	WNPADRDNAIGFYMAVPALAGHKDIGHKLLLPDLQPSNFCLLFDYRLAGDKVKLRV	480
Qy	481	FVKNSNNALAWKTTSEDEKWKTKIQLYQGTDTKSIIFAEARGKGTGEIAVDGVLLV	540
Db	481	FVKNSNNALAWKTTSEDEKWKTKIQLYQGTDTKSIIFAEARGKGTGEIAVDGVLLV	540
Qy	541	SGLCPDLSLSVDD	553
Db	541	SGLCPDLSLSVDD	553

Search completed: May 21, 2004, 12:44:52
Job time : 77.9401 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 21, 2004, 12:28:32 ; Search time 27.2606 Seconds
(without alignments)
1951.314 Million cell updates/sec

Title: US-09-981-649A-24
Perfect score: 3060
Sequence: 1 MPLPWSLALPILLSWVAGF.....VDCVLLVSLCPDLSLSDVD 553

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3016	98.6	558	2 T17324	hypothetical prote
2	410.5	13.4	2907	2 A57278	fibrillin-2 precu
3	409.5	13.4	2918	2 A54105	fibrillin-2 precu
4	387.5	12.7	2871	2 A55567	fibrillin I - bovi
5	385.5	12.6	2871	2 A55624	fibrillin-1 precu
6	383.5	12.5	3002	2 A47221	fibrillin 1 precu
7	382	12.5	1221	2 A49457	fibrillin-2 precu
8	365.5	11.9	1574	2 T13954	MSGF6 protein - ra
9	362.5	11.8	1184	2 A55184	fibulin-2 precu
10	353.5	11.6	1620	2 T27283	hypothetical prote
11	350	11.4	683	2 C36346	fibulin 1 precu
12	348	11.4	685	2 S78040	fibulin, splice fo
13	338	11.0	601	2 B36346	fibulin 1 precu
14	335.5	11.0	2703	1 A24420	notch protein - fr
15	335	10.9	705	2 S34968	fibulin, splice fo
16	332	10.8	2437	2 S42612	transmembrane prot
17	331	10.8	589	2 T43210	fibulin-1D precurs
18	331	10.8	689	2 T42760	fibulin, splice fo
19	330	10.8	712	2 T42990	fibulin 1, splice
20	330	10.8	798	2 T27293	hypothetical prote
21	329	10.8	1712	2 A38261	masking protein pr
22	327.5	10.7	2318	2 S45306	notch 3 protein -
23	327.5	10.7	2321	2 S78549	notch3 protein - h
24	326	10.7	741	2 T46498	hypothetical prote
25	326	10.7	3507	2 T34513	hypothetical prote
26	325.5	10.6	1820	2 A55494	latent transformin
27	325	10.6	2531	2 S18188	notch protein homo
28	322.5	10.5	886	2 A57112	probable hormone r
29	322.5	10.5	2555	2 A40043	notch protein homo

30	322	10.5	2531	2 A46019	notch-1 protein -
31	320	10.5	1394	2 A35626	transforming growt
32	315	10.3	1364	2 T09059	notch4 - mouse
33	311	10.2	810	2 T10756	Nel-homolog protei
34	307.5	10.0	2471	2 A49128	cell-fate determin
35	303.5	9.9	2524	2 A35844	Xotch protein - Af
36	296.5	9.7	1220	2 A56136	jagged protein pre
37	294.5	9.6	2139	2 A35672	crumbs protein - f
38	291.5	9.5	1081	2 T31329	receptor tyrosine
39	291	9.5	387	2 I38449	extracellular prot
40	290	9.5	493	2 JC5621	epidermal growth f
41	289.5	9.5	2352	2 T30201	Notch homolog prot
42	287.5	9.4	722	2 I48324	DELTA-like 1 - mou
43	280	9.2	1106	2 T18739	hypothetical prote
44	279	9.1	1251	2 A57293	latent transformin
45	278	9.1	3623	2 T09456	intrinsic factor-B

ALIGNMENTS

RESULT 1

T17324
hypothetical protein DKFp564P2063.1 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T17324
R;Duesterhoeft, A.; Lauber, J.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, September 1999
A;Reference number: Z18727
A;Accession: T17324
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-558 <DUE>
A;Cross-references: EMBL:AL117610
A;Experimental source: fetal brain; clone DKFp564P2063
C;Genetics:
A;Note: DKFp564P2063.1

Query Match	98.6%	Score	3016;	DB	2;	Length	558;
Best Local Similarity	98.9%	Pred. No.	4.5e-187;				
Matches	547;	Conservative	0;	Mismatches	6;	Indels	0;
Gaps	0;						
Qy	1	MPLPWSLALPILLSWVAGFGNNAARHGLLASARQPGVCHYGTKLACVGRNRSGV	60				
Db	6	MPLPWSLALPILLSWVAGFGNNAARHGLLASARQPGVCHYGTKLACVGRNRSGV	65				
Qy	61	CEATCPGCKFGECVGNPKRCFPGYTGKTCSDVNECGMKPRPCQHRVCVNTGSKYKFC	120				
Db	66	CEATCPGCKFGECVGNPKRCCLPGYTGKTCSDVNECGMKPRPCQHRVCVNTGSKYKFC	125				
Qy	121	LSGHMLMPDATCVNRTCAMINCOYSCDTEEGPQCLCPSSGLRLAPNGDCLDIDECAS	180				
Db	126	LSGHMLMPDATCVNRTCAMINCOYSCDTEEGPQCLCPSSGLRLAPNGDCLDIDECAS	185				
Qy	181	GKVICPNRCVNTGSKYCKHIGLEQVIGRYDCIDINECTMDSHTCSHANCENTQ	240				
Db	186	GKVICPNRCVNTGSKYCKHIGLEQVIGRYDCIDINECTMDSHTCSHANCENTQ	245				
Qy	241	GSKFKCKQGYKNGLRCSAIPENSVEVLRAFGTTIKDRIKLLAHKNSMKKAKIKNVT	300				
Db	246	GSKFKCKQGYKNGLRCSAIPENSVEVLRAFGTTIKDRIKLLAHKNSMKKAKIKNVT	305				
Qy	301	PEPTRTPTKVNLPNPNYBEIVSRGNSHGKKGNEKKEGLEDEKREKALKNDIEER	360				
Db	306	PEPTRTPTKVNLPNPNYBEIVSRGNSHGKKGNEKKEGLEDEKREKALKNDIEER	365				
Qy	361	SLRGDVFPPKVNAGEEGLTLVQRKALTSLKLEKDLNISVDCSFNHCIDWKQDREDDFD	420				
Db	366	SLRGDVFPPKVNAGEEGLTLVQRKALTSLKLEKDLNISVDCSFNHCIDWKQDREDDFD	425				
Qy	421	WNPADRNATGFMVAVPALAGHKDKIRLKLPLDLPQPSNFCLLFDYRLAGDKVGLRV	480				

Db 426 WNPADRDNAIGYMAVPALAGHKDIGRLKLLPDLQPSNFCSLFDYRLAGDKVGLRV 485
QY 481 FVKNSNALANEKTSDEKWKTKIOLYQGTATKSIIFAEKRGKGTGAIANDGVLLV 540
Db 486 FVKNSNALANEKTSDEKWKTKIOLYQGTATKSIIFAEKRGKGTGAIANDGVLLV 545
QY 541 SGLCPDLSLLSVD 553
Db 546 SGLCPDLSLLSVD 558

RESULT 2

A57278
A:Title: fibrillin-2 precursor - mouse
C:Species: Mus musculus (house mouse)
C>Date: 23-Feb-1996 #sequence_revision 23-Feb-1996 #text_change 24-Nov-2003
C:Accession: A57278
J. Zhang, H.; Hu, W.; Ramirez, F.
J. Cell Biol. 129, 1165-1176, 1995
A:Title: Developmental expression of fibrillin genes suggests heterogeneity of extracellular matrix
A:Reference number: A57278; MUID:95263670; PMID:7744963
A:Accession: A57278
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2907 <ZHA>
A:Cross-references: GB:L39790; NID:9762830; PIDN:AAA74908.1; PID:9762831
C:Superfamily: fibrillin; EGF homology
F:1239-1274/Domain: EGF homology <EGF1>
F:1248-2523/Domain: EGF homology <EGF>

Query Match 13.4%; Score 410.5; DB 2; Length 2907;
Best Local Similarity 28.4%; Pred. No. 2.3e-18;
Matches 95; Conservative 37; Mismatches 91; Indels 111; Gaps 13;
QY 38 PGVCHYGT-----KLACCYGW-----RRNSKGVCEATCEPG-CKEGECV---GPNKC 80
Db 1074 PGMCTYGRNRTIGSFKRCNNGFALDMEERNCTDIDECRISDLCSGICVNTPGSPEF 1133
QY 81 RCFPGY-----TGKTCGQ-----D 94
Db 1134 ECFEGYSGFMMKNCMDIDECERNPLLCRGTCVNTGSGFQCDPLGHELSPSREDCVD 1193
QY 95 VNECGMKRPPQCH-RCVNTHGSKYKFCLSGHLMPDAT-CVNSRTCAMIN--CQYSCEDT 150
Db 1194 INECSLDNLCRNGKCVNMIGTYQSCNPGYQATPDQGGCTDIDECMIMNGGCDTQCTNS 1253
QY 151 EGGPQCLPSSGLRLAENGRDCLDIDECASGVIC-----PYNRRCV----- 192
Db 1254 EGSYECSC-SEGYALMPDGRSCADIDECENNPDI CDGGQCTNIPGEYRCLCYDGFMAASMD 1312
QY 193 -----NTFGSYCKCHIGFELQYISGRYDCIDINECTMDSHT 229
Db 1313 MKTCTDVNECDLNPIMCFGECENTKGSFICHQOLGYSVK--KGTTCGCTDVDECEIGAHN 1370
QY 230 CSHHANCFTQSPKCKKQYKNGLRCSAIPES 263
Db 1371 CDHMASCLNIPGSKFCSCREGVNGIKICLDLDE 1404

RESULT 3

A54105
A:Title: fibrillin-2 precursor - human
C:Species: Homo sapiens (man)
C>Date: 09-Sep-1994 #sequence_revision 09-Sep-1994 #text_change 24-Nov-2003
C:Accession: A54105; S17063; S31101
J. Zhang, H.; Apfelloth, S.D.; Hu, W.; Davis, E.C.; Sanguinetti, C.; Bonadio, J.; Mecham, R.
J. Cell Biol. 124, 855-863, 1994
A:Title: Structure and expression of fibrillin-2, a novel microfibrillar component preferentially expressed in cartilage
A:Reference number: A54105; MUID:94165150; PMID:8120105
A:Accession: A54105
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-2918 <ZHA>

A:Cross-references: GB:U03272
R:Lee, B.; Godfrey, M.; Vitale, E.; Hori, H.; Mattei, M.G.; Sarfarazi, M.; Tsipouras, P.
Nature 352, 330-334, 1991
A:Title: Linkage of Marfan syndrome and a phenotypically related disorder to two different genes on chromosome 15
A:Reference number: S17062; MUID:91304567; PMID:1652206
A:Accession: S17063
A:Molecule type: mRNA
A:Residues: 752-1489, 1791, 'GS', 1794-1796, 'QLI', 1922-1923, 'LD', 1926, 'P', 1928 <LEE>
A:Cross-references: EMBL:X62009
R: Milewicz, D.M.
submitted to the EMBL Data Library, December 1992
A:Reference number: S31101
A:Accession: S31101
A:Molecule type: mRNA
A:Residues: 752-1407, 'R', 1409-1489, 1791, 'GS', 1794-1796, 'QLI', 1922-1923, 'LD', 1926, 'P', 19
A:Cross-references: EMBL:X62009
C:Genetics:
A:Gene: GDB:FBN2
A:Cross-references: GDB:128122; OMIM:121050
A:Map position: 5q23-5q31
C:Superfamily: fibrillin; EGF homology
C:Keywords: extracellular protein
F:1-29/Domain: signal sequence #status predicted <SIG>
F:30-2918/Product: fibrillin-2 #status predicted <MAT>
F:1245-1280/Domain: EGF homology <EGF1>
F:1970-2013/Domain: EGF homology <EGF>

Query Match 13.4%; Score 409.5; DB 2; Length 2918;
Best Local Similarity 28.6%; Pred. No. 2.7e-18;
Matches 96; Conservative 37; Mismatches 92; Indels 111; Gaps 13;

QY 38 PGVCHYGT-----KLACCYGW-----RRNSKGVCEATCEPG-CKEGECV---GPNKC 80
Db 1080 PGMCTYGRNRTIGSFKRCNNGFALDMEERNCTDIDECRISDLCSGICVNTPGSPEF 1139
QY 81 RCFPGY-----TGKTCGQ-----D 94
Db 1140 ECFEGYSGFMMKNCMDIDECERNPLLCRGTCVNTGSGFQCDPLGHELSPSREDCVD 1199
QY 95 VNECGMKRPPQCH-RCVNTHGSKYKFCLSGHLMPDAT-CVNSRTCAMIN--CQYSCEDT 150
Db 1200 INECSLDNLCRNGKCVNMIGTYQSCNPGYQATPDQGGCTDIDECMIMNGGCDTQCTNS 1259
QY 151 EGGPQCLPSSGLRLAENGRDCLDIDECASGVIC-----PYNRRCV----- 192
Db 1260 EGSYECSC-SEGYALMPDGRSCADIDECENNPDI CDGGQCTNIPGEYRCLCYDGFMAASMD 1318
QY 193 -----NTFGSYCKCHIGFELQYISGRYDCIDINECTMDSHT 229
Db 1319 MKTCTDVNECDLNPIMCFGECENTKGSFICHQOLGYSVK--KGTTCGCTDVDECEIGAHN 1376
QY 230 CSHHANCFTQSPKCKKQYKNGLRCSAIPENS 265
Db 1377 CDHMASCLNIPGSKFCSCREGVNGIKICLDLDECS 1412

RESULT 4

A55567
A:Title: fibrillin I - bovine
C:Species: Bos primigenius taurus (cattle)
C>Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 24-Nov-2003
C:Accession: A55567
R: Tillet, D.J.; Li, L.; Potter, K.A.; Womack, J.; Byers, P.H.
Genomics 23, 480-485, 1994
A:Title: Sequence of the coding region of the bovine fibrillin cDNA and localization to chromosome 12
A:Reference number: A55567; MUID:95137597; PMID:7835900
A:Accession: A55567
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-2871 <TIL>
A:Cross-references: GB:L28748; NID:9508427; PIDN:AAA74122.1; PID:9508428
C:Superfamily: fibrillin; EGF homology
F:1201-1236/Domain: EGF homology <EGF>

Query Match 12.7%; Score 387.5; DB 2; Length 2871;
Best Local Similarity 30.6%; Pred. No. 6.9e-17;
Matches 89; Conservative 34; Mismatches 91; Indels 77; Gaps 10;
QY 49 CCYGMRRNSKGVCEATCEPG-----CKFGECV---GPNKCR 82
DB 1081 CGRGQCQVTPGDFECKDEGYESGFMNMKNMCMIDBCQDRDPLLRCRGVCLNTEGSRCEC 1140
QY 83 FPGY---TGKTCQDVNCGMKRPPCOH-RCVNTGSKYKFCCLSGHMLPDAI-CVNSRT 137
DB 1141 PPGHQLAPNISACIDINECELSAHLCPHGRCVNLGKYQACNPGVSHYTPDRFLFCVDIDE 1200
QY 138 CAMIN--COYSCEDTBEGPCLCFSSGLRLAPNGRDCLDIDECASGVIC-----PY 187
DB 1201 CSIMNGCETFCNTEGSEYECSC-QPGFALMPQSRCTDIDECEDNPNICDGGQCTNIFG 1259
QY 188 NRRCV-----NTFGSYKCKHIGFELQYISGR 214
DB 1260 EYRCLCYDGFMASEDMMKTCVDVNECDLNPICLSGTCENTKGSFICHDMGYSGK--KGK 1317
QY 215 YDCIDINECTWDSHTCSHANCENTQGSFKCKQKQVKGNGLRCSAIPENS 265
DB 1318 TGCTDINECEIGHANCDRAVCTNTAGSFKCSCPGWIGDGIKCTDLDECS 1368
RESULT 5
A55624
fibrillin-1 precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 23-Mar-1995 #sequence_revision 23-Mar-1995 #text_change 24-Nov-2003
C:Accession: A55624
R:Yin, W.; Smiley, E.; Germiller, J.; Sanguinetti, C.; Lawton, T.; Pereira, L.; Ramirez, J. Biol. Chem. 270, 1798-1806, 1995
A:Title: Primary structure and developmental expression of Fbn-1, the mouse fibrillin gene
A:Reference number: A55624; MUID:95130561; PMID:7829516
A:Accession: A55624
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-2871 <N>
A:Cross-references: GB:L29454; NID:9575509; PIDN:AAA56840.1; PID:9575510
C:Genetics:
A:Gene: Fbn-1
C:Superfamily: fibrillin; EGF homology
F:1201-1236/Domain: EGF homology <EGF>
Query Match 12.6%; Score 385.5; DB 2; Length 2871;
Best Local Similarity 32.2%; Pred. No. 9.3e-17;
Matches 86; Conservative 34; Mismatches 82; Indels 65; Gaps 11;
QY 39 GVCHYGTKLACCYGWRNRNSKGVCEATCEPG-----CKFGECV- 75
DB 1127 GICH-----NTGTYRCECFPHGQLSPNISACIDINECELSANLCPHGRCVN 1173
QY 76 --GNKCRCPFGY--TGKTCQDVNCGMKRPPCQHRVCNTHGSKYKFCCLSGHMLPDA 130
DB 1174 LIGKYQACNPGVHPTDRFLFCVDIDECSIMNGGCETFCNTEGSEYECSCQPGFALMPD- 1232
QY 131 TCVNSRFCAMINCOYSCEDTIE--EGPQ-----CLCPSSGLRLAPNGRDCLDIDEC 178
DB 1233 ----QRSCDID----QCEDNPNICDGGQCTNIPGEYRCLC-YDGFMASEDMMKTCVDVNEC 1284
QY 179 ASGVICPYNRRVCNVTGSKYKCHIGFELQYISGRYDCIDINECTWDSHTCSHANCFN 238
DB 1285 DLNPNIC-LSGTCENTKGSFICHDMGYSGK--KGKTCIDINECEIGHANCDRAVCTN 1341
QY 239 TQGSFKCKCKQKQVKGNGLRCSAIPENS 265
DB 1342 TAGSFKCSCPGWIGDGIKCTDLDECS 1368
RESULT 6
A47221

fibrillin 1 precursor - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 02-Jun-1995 #sequence_revision 25-Apr-1997 #text_change 24-Nov-2003
C:Accession: A47221; I54355; S17064; I59574; S17062; S62111; A34198
R:Corson, G.M.; Chalberg, S.C.; Dietz, H.C.; Charbonneau, N.L.; Sakai, L.Y. Genomics 17, 476-484, 1993
A:Title: Fibrillin binds calcium and is coded by cDNAs that reveal a multidomain structure
A:Reference number: A47221; MUID:94010947; PMID:7691719
A:Accession: A47221
A:Molecule type: mRNA
A:Residues: 1-337, 'T', 339-1029 <OR>
A:Cross-references: GB:X63556
R:Pereira, L.V.; D'Alessio, M.; Ramirez, F.; Lynch, J.; Sykes, B.; Pangilinan, T.; Bonafant-Hopfe, E. Hum. Mol. Genet. 2, 961-968, 1993
A:Title: Genomic organization of the sequence coding for fibrillin, the defective gene in Marfan syndrome
A:Reference number: I54355; MUID:93372860; PMID:8364578
A:Accession: I54355
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 132-3002 <PER>
A:Cross-references: GB:L13923; NID:9306745; PIDN:AAB02036.1; PID:9306746
R:Maslen, C.L.; Corson, G.M.; Maddox, B.K.; Glanville, R.W.; Sakai, L.Y. Nature 352, 334-337, 1991
A:Title: Partial sequence of a candidate gene for the Marfan syndrome.
A:Reference number: S17064; MUID:91304568; PMID:1852207
A:Accession: S17064
A:Molecule type: mRNA
A:Residues: 1030-3002 <MAS>
A:Cross-references: EMBL:X63556
R:Dietz, H.C.; Valle, D.; Francomano, C.A.; Kendzior, R.J. Science 259, 680-683, 1993
A:Title: The skipping of constitutive exons in vivo induced by nonsense mutations.
A:Reference number: I59574; MUID:93157831; PMID:8430317
A:Accession: I59574
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 2217-2288, 'I', 2290-2325 <RES>
A:Cross-references: GB:S54426; NID:9264860; PIDN:AAB25244.1; PID:9264861
R:Lee, B.; Godfrey, M.; Vitale, E.; Hori, H.; Mattei, M.G.; Sarrafzadi, M.; Tsipouras, P. Nature 352, 330-334, 1991
A:Title: Linkage of Marfan syndrome and a phenotypically related disorder to two different chromosomes
A:Reference number: S17062; MUID:91304567; PMID:1852206
A:Accession: S17062
A:Molecule type: protein
A:Residues: 1166-1176, 'X', 1178-1180, 'D', 1182-1185 <LEE2>
R:Maddox, B.K.; Sakai, L.Y.; Keene, D.R.; Glanville, R.W. J. Biol. Chem. 264, 21381-21385, 1989
A:Title: Connective tissue microfibrils. Isolation and characterization of three large fragments
A:Reference number: A34198; MUID:90078246; PMID:2512293
A:Accession: A34198
A:Molecule type: protein
A:Residues: 585-575, 1890-1892, 'I', 1894-1900 <MAD>
C:Comment: Fibrillin is a major component of elastin-associated microfibrils.
C:Genetics:
A:Gene: GDB:FBNI
A:Cross-references: GDB:127115; OMIM:134797; OMIM:154700
A:Map position: 15q21.1-15q21.1
A:Introns: 2236/1; 2259/1; 2297/1
C:Superfamily: fibrillin; EGF homology
C:Keywords: alternative splicing; calcium binding; extracellular matrix; glycoprotein; N-linked glycosylation
F:1-3002/Product: fibrillin (5'-region exon A splice form) #status predicted
F:132-3002/Product: fibrillin (5'-region exon C splice form) #status predicted <MATC>
F:1332-1367/Domain: EGF homology <EGF>
F:1457-1492/Domain: EGF homology <EGF2>
F:2262-2285/Domain: EGF homology <EGF1>
Query Match 12.5%; Score 383.5; DB 2; Length 3002;
Best Local Similarity 32.6%; Pred. No. 1.3e-16;
Matches 87; Conservative 33; Mismatches 82; Indels 65; Gaps 11;

C>Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 08-Sep-2002
C/Accession: A49457; S74095
C/Accession: A49457; S74095
R;Pan, T.C.; Sasaki, T.; Zhang, R.Z.; Faessler, R.; Timpl, R.; Chu, M.L.
J. Cell Biol. 123, 1269-1277, 1993
A>Title: Structure and expression of fibulin-2, a novel extracellular matrix protein with
A/Reference number: A49457; MUID:94064787; PMID:8245130
A/Accession: A49457
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-1221 <PAN>
A/Cross-references: NID:G437046; PIDN:CAAS3040.1; PID:G437047
R;Sasaki, T.; Mann, K.; Murphy, G.; Chu, M.L.; Timpl, R.
Eur. J. Biochem. 240, 427-434, 1996
A>Title: Different susceptibilities of fibulin-1 and fibulin-2 to cleavage by matrix metalloproteinases
A/Reference number: S74094; MUID:96439073; PMID:8841408
A/Accession: S74095
A/Molecule type: protein
A/Residues: 236-238,'X',240-247;260-275;336-344,'L',346-361;405-426;566-568,'EM',569-589
C/Superfamily: fibulin-2; EGF homology; von Willebrand factor type C repeat homology
C/Keywords: calcium binding; duplication; extracellular matrix; glycoprotein; homotrimer
F:942-978/Domain: EGF homology <EGF>

```

Query Match          11.9%; Score 365.5; DB 2; Length 1574;
Best Local Similarity 36.8%; Pred.No. 9.9e-16;
Matches 93; Conservative 25; Mismatches 96; Indels 39; Gaps 13;

QY      34  SARQPGVCHYGTKLACCYGWR-----NSKGVCEATCEPGCKFGCECVGN 78
Db      92  SYRQVYATEARTVFRCCPGWSQKPGQSGCLSDVDDECASANGCGPC---C---NTVGGF 145

QY      79  KRCFCFGY----TGKTSQDVNECGMKPRPCQHRHCYVNTGHSYKFCFLSGHMLMPDA-TCV 133
Db      146  YCRCPGYQLQDGKTC-QVDVDECRAHNGGCGHRCVNTPGSYLCECKPGFRLHTDGRITCL 204

QY      134  NSRTCAMIN--CQYSC-EDTEEGPOCLCPSSGLFLAPNGRDCLDIDECASGKVICPNYR 190
Db      205  A1SSCTLNGGCGHCQCVLTQTHRCQCRPQ-YOLQDGRRCVRRSPCAENGGGMH--I 261

QY      191  CVNTFGSYCKCHIGFELQVISGRYDCIDINECTMDSHTCSHHANCFNTQGSFKCKKQG 250
Db      262  CQELRGLAHGCHPGYQL--AADKTCEDVDECALGLAQCAH--GCLNTQGSFKVCVCHAG 317

QY      251  YK--GNGLRCSAI 261
Db      318  YELGADGRQCVRI 330

RESULT 9
A55184
fibulin-2 precursor - human
N:Alternate names: protein DKFp566A1519.1
C:Species: Homo sapiens (man)
C:Date: 27-Jan-1995 #sequence_revision 27-Jan-1995 #text_change 08-Sep-2002
C:Accession: A55184; T08744
R:Zhang, R.Z.; Pan, T.C.; Zhang, Z.Y.; Mattei, M.G.; Timpl, R.; Chu, M.L.
Genomics 22, 425-430, 1994
A:Title: Fibulin-2 (FBLN2): human cDNA sequence, mRNA expression, and mapping of the gene
A:Reference number: A55184; MUID:95104855; PMID:7806230
A:Accession: A55184
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1184 <ZHA>
A:Cross-references: GB:X82494; NID:G575232; PIDN:CRA57876.1; PID:G575233
R:Wambutt, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, March 1999
A:Reference number: Z16471
A:Accession: T08744
A:Molecule type: mRNA

```

A;Residues: 656-719, 'QDECLMGADHCSRQFCVNTLGSFYCNHTVLQADGYILNAHRKQVD', 720-853, 'T', 855-111
A;Cross-references: EMBL:AL050095
A;Experimental source: adult uterus; clone DKFp586A1519
C;Genetics:
A;Gene: GDB:FBIN2
A;Cross-references: GDB:293037; OMIM:135821
A;Map position: 3p25-3p24
A;Note: DKFp586A1519.1
C;Superfamily: fibulin-2; EGF homology; von Willebrand factor type C repeat homology
C;Keywords: alternative splicing; extracellular matrix
F;1-27/Domain: signal sequence #status predicted <SIG>
F;28-1184/Product: fibulin-2 protein #status predicted <MAT>
F;905-941/Domain: EGF homology <EGF>

Query Match 11.8%; Score 362.5; DB 2; Length 1184;
Best Local Similarity 30.7%; Pred. No. 1.2e-15;
Matches 107; Conservative 35; Mismatches 149; Indels 57; Gaps 17;

QY 21 GNAASARHHCLLASARQPGVCHYGT-----KLACCYWRNSKG---VCEATCEPG 68
Db 806 GNCVDINECTSLSEPCRPFGSCINTVGSYTCQRNPICARGYHASDDGAKCVDVNECETG 865
QY 69 ---CKTGEVC---GNPKRCFFGYT----GKTCSDQDVNECGMKP-RPCQHRVCWNTGSGY 116
Db 866 VHRGCGQVCHNLPGSYRCDKAGFQDAFGRGCI-DVNECWASPGELCQHTCENTILGSGY 924
QY 117 KFCFLSGHMLMPDA-TCVNSRTCAMINCQVSCDETEGQCCLPSSGLRLAPNGRDCLDI 175
Db 925 RSCASGFLLAADGKRCEDVNECAQCSQECANIYGSQCYC-RQGYLAEDGHTCTDI 983
QY 176 DECSAGK-VICPNRRVCNTFGSYCKCHIGFELYISGRYDCIDINECTMDSHH 234
Db 984 DECAQAGILICTP-RCLNVPSGYQACPEQGYTMTANGR-SCKDVDECALGTHNCSEAE 1040
QY 235 NCFNTQGSFKC---KCKQGY-KGNGLRCSAIPENSVKVEVLRAPTGKIKLLAHKNSM 290
Db 1041 TCHNIQSFCLPECPNPVQVSKYCETTTCHDFLECONSAPRI----THYQLNQTGL 1097
QY 291 KKKAKIKNVTPETRT-PTPKVNLQPNFVEIIVSRGNSHGKGNNE 337
Db 1098 LVPAHIFRIGPAFPAFTGDTIALNI-----IKGNNEE 1127

RESULT 10
T27283
hypothetical protein Y64G10A.f - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T27283
R;Ainscough, R.
submitted to the EMBL Data Library, September 1999
A;Reference number: Z20336
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Accession: T27283
A;Molecule type: DNA
A;Residues: 1-1620 <MIL>
A;Cross-references: EMBL:AL110498; NID:el542303; PIDN:CA854471.1; CESP:Y64G10A.f
A;Experimental source: clone Y64G10A
C;Genetics:
A;Gene: CESP:Y64G10A.f
A;Introns: 77/1; 116/1; 198/1; 282/1; 365/1; 425/1; 466/1; 548/1; 559/1; 601/1; 625/1; 7

Query Match 11.6%; Score 353.5; DB 2; Length 1620;
Best Local Similarity 33.9%; Pred. No. 6e-15;
Matches 82; Conservative 38; Mismatches 81; Indels 41; Gaps 15;

QY 40 VCHGTKLACCYWRNSKGVC-----ATCE-----PGCKFGEVC---GPNK 79
Db 83 LCHNGG--TCVPSEHNDNEQVCEPVGFTGAKCOYDANECAVNGGCEH-ECVNTIGTYV 139
QY 80 CRCPFGY---TGKTCSDQDVNECGMKRPPQHRVCWNTFGSYKFCFLSGHMLMPDA-TCVN 134
Db 140 CRCPFGELSGDGNTCs-DIDECAVNGGSDRCVNSPGGFRCDPDLYLHADGRTCGK 198

QY 135 SRTCAMIN--COYSCSDTEGP--QCLCPSSGLRLAPNGRDCLDIDECASGKVICPNRR 190
Db 199 VTSCSTDNGGCEHCENDSNGBFYRCRC-RVGFKLSENKRSCQVDFPCFDNKGCGQH-- 255
QY 191 CVNTEGSSYYCKCHIGFELYISGRYDCIDINECTMDSHHANCFTQGSFKCKCKQG 250
Db 256 CYNHGRAOCCYPPGPHLSY--DRRSCVDIDECA-KNNGCEHF--CENVKGYTRCKCREG 310
QY 251 YK 252
Db 311 YQ 312

RESULT 11
C36346
fibulin 1 precursor, splice form C - human
N;Alternate names: fibulin C
N;Contains: fibulin 1 splice form A; fibulin 1 splice form C
C;Species: Homo sapiens (man)
C;Date: 19-Apr-1991 #sequence_revision 19-Apr-1991 #text_change 02-Aug-2002
C;Accession: C36346; A36346; A32826
R;Argaves, W.S.; Tran, H.; Burgess, W.H.; Dickerson, K.
J. Cell Biol. 111, 3155-3164, 1990
A;Title: Fibulin is an extracellular matrix and plasma glycoprotein with repeated domain
A;Reference number: A36346; MUID:31100426; PMID:2269669
A;Accession: C36346
A;Molecule type: mRNA
A;Residues: 1-683 <ARG>
A;Cross-references: GB:X53743; NID:g31418; PIDN:CAA37772.1; PID:g31419
A;Accession: A36346
A;Molecule type: mRNA
A;Residues: 1-566 <AR2>
A;Cross-references: GB:X53741; NID:g31414; PIDN:CAA37770.1; PID:g31415
R;Argaves, W.S.; Dickerson, K.; Burgess, W.H.; Ruoslahti, E.
Cell 58, 623-629, 1989
A;Title: Fibulin, a novel protein that interacts with the fibronectin receptor beta-sub
A;Reference number: A32826; MUID:89354537; PMID:2527614
A;Accession: A32826
A;Molecule type: protein
A;Residues: 30-35,'SX',38-40,'SH',43-44 <AR3>
C;Genetics:
A;Gene: GDB:FBIN1; FBLN
A;Cross-references: GDB:278285; OMIM:135820
A;Map position: 22q13.3-22q13.3
C;Superfamily: fibulin-1; EGF homology
C;Keywords: alternative splicing; glycoprotein
F;1-29/Domain: signal sequence #status predicted <SIG>
F;30-683/Product: fibulin 1 splice form C #status predicted <MAT>
F;180-214/Domain: EGF homology <EGF>
F;485-523/Domain: EGF homology <EGF>
F;98,535,539/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 11.4%; Score 350; DB 2; Length 683;
Best Local Similarity 31.9%; Pred. No. 4.3e-15;
Matches 100; Conservative 37; Mismatches 120; Indels 56; Gaps 18;

QY 49 CCYWRNSKGV-----CEATCEPGCKFGEVC---GNPKRCFFGY-----TGKTCSD 94
Db 341 CGRGYHLEEGTRCVDVDECAPPAEPCGKHRCVNSPGSFRCECKTGYFDGISRMC-VD 399
QY 95 VNECGMKP-RPCQHRVCWNTFGSYKFCFLSGHMLMPDA-TCVNSRTCAMINCOYSCSDTEE 152
Db 400 VNECQYRGLCHKCENTILGSCSVGFLVSDVGRSCDINEGSSFCSCQECANVYG 459
QY 153 GPCLCPSSGLRLAP-NGRDCLDIDEC--SGKVICPNRRVCWNTFGSYKCK-HIGFEL 208
Db 460 SYQCYC-RRGYQLSDVDGVTCEIDECALPTGCHISY--RCINIFGFCQSCPSGSGYL 516
QY 209 QYISGRYDCIDINECTMDSHHANCFTQGSFKC---KCKQGYKGN-GLRCSAIPEN 264
Db 517 A-PNGR-NQDIDECVTHGNCINETCFNIQAGFLAFCECPENYRRSAATCERLPCH 574

A;Accession number: 009358
A;Accession: S09358
A;Molecule type: DNA
A;Residues: 2505-2551, 'QQQ', 2552-2576, 'E', 2578-2604 <TAU>
R:Wharton, K.A.; Vedvobnick, B.; Finnerty, V.G.; Attavanis-Taakonas, S.
Cell 40, 55-62, 1985
A;Title: opa: a novel family of transcribed repeats shared by the Notch locus and other
A;Reference number: A05267; MUID:85099329; PMID:2981631
A;Accession: A05267

A:Molecule type: DNA
A:Residues: 2504-2576, 'E', 2578-2611 <WHA2>
C:Genetics:
A:Gene: notch; opa
A:Cross-references: FlyBase:FBgn0004647
A:Map position: 8.96-9.36
A:Introns: 53/3; 84/3; 171/3; 240/3; 283/3; 2333/3; 2436/3; 2588/3
C:Superfamily: notch protein; ankyrin repeat homology; EGF homology
C:Keywords: differentiation; tandem repeat; transmembrane protein
F:27-43/Domain: transmembrane #status predicted <TM1>
F:297-328/Domain: EGF homology <EGX1>
F:530-561/Domain: EGF homology <EGF1>
F:568-595/Domain: EGF homology <EGF>
F:988-1019/Domain: EGF homology <EGX2>
F:1084-1095/Domain: EGF homology <EGF3>
F:1187-1218/Domain: EGF homology <EGX3>
F:1746-1762/Domain: transmembrane #status predicted <TM2>
F:1950-1982/Domain: ankyrin repeat homology <AN1>
F:1983-2015/Domain: ankyrin repeat homology <AN2>
F:1988-2004/Domain: transmembrane #status predicted <TM3>
F:2017-2049/Domain: ankyrin repeat homology <AN3>
F:2083-2115/Domain: ankyrin repeat homology <AN4>
F:2083-2115/Domain: ankyrin repeat homology <AN5>
F:2538-2568/Region: glutamine-rich
F:2538-2568/Domain: neurogenic repetitive element #status predicted <OPA>

Query Match 11.0%; Score 335.5; DB 1; Length 2703;
Best Local Similarity 26.4%; Pred. No. 1.5e-13;
Matches 96; Conservative 34; Mismatches 106; Indels 127; Gaps 17;

Qy 39 GVCHYGT----KLACCYGW---RRNSKGVCEAT-CEPCCKFGCEVGNPK--CRCFPGYTG 88
Db 152 GTCQLKLEETACANGYTGERCETKMLCASSPCRNATCTALAGSSSFTCSPPGFTG 211

Qy 89 KTCSDQVNECGMKPRPCQ--RCVNTGSGYKCFCLSG-----HMLMPDATCVNSRT 137
Db 212 DTCSDYIEEC--QSNPKYGGTCVNTGSGYCMCFPTGYTKDCTKYKPCSPQNGGI 269

Qy 138 CAMINCOYSQSDTE--EGPQC-----LCPSG-----LRAPN--GRDCL 173
Db 270 CRSNGLSYECKCPKGFEGKNEQNYDCLGLHCQNGGTCIDGISDYTCRCPNFTGRFCQ 329

Qy 174 -DIDECAS-----FELQYISGRYDCI-----D 219
Db 330 DDVDECAQRDPVPCQNGATCINTHGSYSCICVNGWAGLDCSNNTDDCKQAACFYGATCID 389

Qy 194 TFGSYCKCHIG-----FELQYISGRYDCI-----D 219
Db 390 GVGsfYQCCTKGTGLLLCHLDDACTSNPCHADAICDTSPINGSYACSCATGYKGVDCSED 449

Qy 220 INECTMDSHTCSHANCENTOGSFCKKQYKG-----NGLRCSAIPENSVEVLRAPG 274
Db 450 IDECDQGS-PCENHNGICVNTPGSYRCNSQSGFTGPRCTNINECESHPQNEGSLDDPG 508

Qy 275 TIK 277
Db 509 TFR 511

RESULT 15
S34968
fibulin, splice form D precursor - mouse
N:Alternate names: basement-membrane protein BM-90; calcium-binding protein BM-90
C:Species: Mus musculus (house mouse)
C:Date: 10-Dec-1993 #sequence revision 10-Nov-1995 #text_change 02-Aug-2002
C:Accession: S34968; S36441; S13814
R:Pan, T.C.; Kluge, M.; Zhang, R.Z.; Mayer, U.; Timpl, R.; Chu, M.L.
Eur. J. Biochem. 215, 733-740, 1993
A:Title: Sequence of extracellular mouse protein BM-90/fibulin and its calcium-dependent
A:Reference number: S34968; MUID:93358897; PMID:8354280
A:Accession: S34968
A:Molecule type: mRNA
A:Residues: 1-705 <PAN>

R:Pan, T.C.; Kluge, M.; Zhang, R.Z.; Mayer, U.; Timpl, R.; Chu, M.L.
submitted to the EMBL Data Library, January 1993
A:Description: Sequence of extracellular mouse protein BM-90/fibulin and its calcium-de
A:Reference number: S36440
A:Accession: S36441
A:Molecule type: mRNA
A:Residues: 1-39, P', 41-705 <PAW>
A:Cross-references: EMBL:U0854; NID:G396820; PID:CAAS0207.1; PID:G396821
A:Experimental source: cell-line F9 teratocarcinoma
R:Kluge, M.; Mann, K.; Dziadek, M.; Timpl, R.
Eur. J. Biochem. 193, 651-659, 1990
A:Title: Characterization of a novel calcium-binding 90-kDa glycoprotein (BM-90) shared
A:Reference number: S13814; MUID:91065369; PMID:2249686
A:Accession: S13814

A:Molecule type: protein
A:Residues: 28;31-49,'X',110-117;231-240,'X',242-243;339-362,'S',364-387;434
C:Superfamily: fibulin-1; EGF homology
C:Keywords: alternative splicing; basement membrane; calcium binding; extracellular mat
F:1-29/Domain: signal sequence #status predicted <SIG>
F:30-705/Product: fibulin, splice form D #status predicted <MAT>
F:98,537,541/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 10.9%; Score 335; DB 2; Length 705;
Best Local Similarity 34.1%; Pred. No. 4.1e-14;
Matches 86; Conservative 31; Mismatches 87; Indels 48; Gaps 15;

Qy 46 KLACCYGWRRNSKGVCEAT-CEPCCKFGCEVGNPKCR-----EATCEPGCKFGCEVGNPKCR-----CFPGY----T 87
Db 293 KLQCKSGFIQDALGNCLIDINECLISAPCPVGTCTINTEGTYTCOKNVPNCGRGYHLNEE 352

Qy 88 GKTCSDQVNECGMKPRPC--QHRCVNTHGSGYKCFCLSGHMLMPDATCVNSRTCAVIN--- 142
Db 353 GTRC-VDVDECAAPPAPPCGKGHHCLNSPGRCECKAGFYF--DGI---SRTCVDINECQ 406

Qy 143 -----COYSCDETEGPQCLCPSSGLRLAPNGRDCLDIDECASGVKICPNRRRCVNTFG 196
Db 407 RYGRCLCGHKCENTPGSFHCSC--SAGFRLSVDRGSRCDVNECLNS-----PCSOECANVIG 461

Qy 197 SYCKCHIGFELQYISGRYDCIDINECTMDS--HTCSHANCENTOGSFCKK--KQYK-- 252
Db 462 SYQCYCHRGVQLSDVDG-VTCEDIDECALPTGGHICSY--RCINIPGSFQSCSPSSGYRL 518

Qy 253 -GNGLRCSAIP 263
Db 519 APNGRNCQDIDE 530

Search completed: May 21, 2004, 12:33:56

Job time : 29.2606 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 21, 2004, 12:27:01 ; Search time 19.4718 Seconds
(without alignments)
1478.792 Million cell updates/sec

Title: US-09-981-649A-24
Perfect score: 3060
Sequence: 1 MPLPWSLALPLLSWAGGF.....VDGVLLVSLGCLPDSLLSVDD 553

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	410.5	13.4	2907	1	FBN2_MOUSE
2	409.5	13.4	2911	1	FBN2_HUMAN
3	387.5	12.7	2871	1	FBN1_BOVIN
4	385.5	12.6	2871	1	FBN1_MOUSE
5	385.5	12.6	2871	1	FBN1_PIG
6	383.5	12.5	2871	1	FBN1_HUMAN
7	382	12.5	1221	1	FBL2_MOUSE
8	362.5	11.8	1184	1	FBL2_HUMAN
9	352.5	11.5	598	1	FBL1_CERAE
10	352	11.5	704	1	FBL1_CHICK
11	339	11.1	956	1	MTN2_MOUSE
12	338	11.0	703	1	FBL1_HUMAN
13	335.5	11.0	2703	1	NOTC_DROME
14	335	10.9	705	1	FBL1_MOUSE
15	332	10.8	1389	1	LTBS_MOUSE
16	332	10.8	1713	1	LTBL_MOUSE
17	332	10.8	2437	1	NTC1_BRARE
18	331	10.8	956	1	MTN2_HUMAN
19	330	10.8	798	1	FBL1_CABEL
20	329.5	10.8	681	1	FBL1_BRARE
21	329.5	10.8	2319	1	NTC3_RAT
22	329	10.8	1712	1	LTBL_RAT
23	327.5	10.7	2318	1	NTC3_MOUSE
24	327.5	10.7	2321	1	NTC3_HUMAN
25	325	10.6	2531	1	NTC1_RAT
26	322.5	10.5	886	1	EMRI_HUMAN
27	322.5	10.5	2556	1	NTC1_HUMAN
28	322	10.5	2531	1	NTC1_MOUSE
29	321	10.5	1213	1	JAG3_BRARE
30	320	10.5	1394	1	LTBS_HUMAN
31	320	10.5	1595	1	LTBL_HUMAN
32	315.5	10.3	443	1	FBL4_MOUSE
33	315	10.3	1964	1	NTC4_MOUSE

RESULT 1
FBN2_MOUSE
ID FBN2_MOUSE STANDARD; PRT; 2907 AA.
AC Q61555; O63957;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Fibrillin 2 precursor.
GN FBN2 OR FBN-2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95263670; PubMed=7744963;
RA Zhang H., Hu W., Ramirez F.;
RT "Developmental expression of fibrillin genes suggests heterogeneity of extracellular microfibrils.";
RL J. Cell Biol. 129:1165-1176(1995).
RN [2]
RP SEQUENCE OF 210-317 FROM N.A.
RX MEDLINE=94140368; PubMed=8307578;
RA Li X., Pereira L., Zhang H., Sanguinetti C., Ramirez F., Bonadio J., Francke U.;
RT "Fibrillin genes map to regions of conserved mouse/human synteny on mouse chromosomes 2 and 18.";
RL Genomics 18:667-672(1993).
CC -I- FUNCTION: Structural component of connective tissue microfibrils that binds calcium. Fibrillin-2-containing microfibrils regulate the early process of elastic fiber assembly.
CC -I- SIMILARITY: Contains 47 EGF-like domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L39790; AAA74908.1; -.
CC EMBL; S69359; AAC60685.1; -.
CC PIR; A57278; A57278.
CC HSSP; P35555; 1EMN.
CC MGD; MGI:95490; Fbn2.
CC GO; GO:0030326; P.limb morphogenesis; IMP.
CC InterPro; IPR000152; Abx_hydroxyl_S.
CC InterPro; IPR001881; EGF Ca.
CC InterPro; IPR001438; EGF II.
CC InterPro; IPR006209; EGF-like.
CC InterPro; IPR002212; Fibril-assoc.
CC Pfam; PF00008; EGF; 44.
CC PRINTS; PR00083; TB; 9.
CC PRINTS; PR00010; EGFBL00D.

34	313	10.2	443	1	FBL4_HUMAN
35	313	10.2	448	1	FBL5_HUMAN
36	311.5	10.2	2524	1	NOTC_XENLA
37	311	10.2	448	1	FBL5_RAT
38	311	10.2	810	1	NEL1_RAT
39	311	10.2	835	1	CD97_HUMAN
40	310.5	10.1	810	1	NEL1_HUMAN
41	309.5	10.1	448	1	FBL5_MOUSE
42	308	10.1	443	1	FBL4_CRIGR
43	307.5	10.0	2471	1	NTC2_RAT
44	300	9.8	2470	1	NTC2_MOUSE
45	298.5	9.8	1218	1	JAG1_MOUSE

O95967 homo sapien
O9ubx5 homo sapien
P21783 xenopus lae
Q9vvh8 rattus norv
Q62919 rattus norv
P48960 homo sapien
Q92832 homo sapien
Q9vvh9 mus musculu
O55058 cricetus
Q9q330 rattus norv
O35516 mus musculu
Q9qxx0 mus musculu

ALIGNMENTS

DR SMART, SM00179; EGF_CA; 43.
 DR PROSITE; PS00010; ASX HYDROXYL; 43.
 DR PROSITE; PS00022; EGF_1; 2.
 DR PROSITE; PS01186; EGF_2; 36.
 DR PROSITE; PS00026; EGF_3; 45.
 DR PROSITE; PS01187; EGF_CA; 43.
 KW Extracellular matrix; Calcium-binding; Glycoprotein; EGF-like domain;
 KW Repeat; signal; Multigene family.
 FT SIGNAL 1 28 POTENTIAL.
 FT CHAIN 29 2907 FIBRILLIN 2.
 FT DOMAIN 111 142 EGF-LIKE 1.
 FT DOMAIN 145 176 EGF-LIKE 2.
 FT DOMAIN 176 208 EGF-LIKE 3.
 FT DOMAIN 276 317 EGF-LIKE 4.
 FT DOMAIN 318 359 EGF-LIKE 5, CALCIUM-BINDING.
 FT REPEAT 360 426 TGFBBP 1.
 FT DOMAIN 487 527 EGF-LIKE 6.
 FT DOMAIN 528 567 EGF-LIKE 7, CALCIUM-BINDING.
 FT DOMAIN 568 609 EGF-LIKE 8, CALCIUM-BINDING.
 FT DOMAIN 610 650 EGF-LIKE 9, CALCIUM-BINDING.
 FT DOMAIN 651 691 EGF-LIKE 10, CALCIUM-BINDING.
 FT REPEAT 692 760 TGFBBP 2.
 FT DOMAIN 761 802 EGF-LIKE 11, CALCIUM-BINDING.
 FT DOMAIN 803 844 EGF-LIKE 12, CALCIUM-BINDING.
 FT DOMAIN 845 883 EGF-LIKE 13, CALCIUM-BINDING.
 FT DOMAIN 948 989 EGF-LIKE 14, CALCIUM-BINDING.
 FT REPEAT 990 1065 TGFBBP 3.
 FT DOMAIN 1066 1107 EGF-LIKE 15, CALCIUM-BINDING.
 FT DOMAIN 1108 1150 EGF-LIKE 16, CALCIUM-BINDING.
 FT DOMAIN 1151 1192 EGF-LIKE 17, CALCIUM-BINDING.
 FT DOMAIN 1193 1234 EGF-LIKE 18, CALCIUM-BINDING.
 FT DOMAIN 1235 1275 EGF-LIKE 19, CALCIUM-BINDING.
 FT DOMAIN 1276 1317 EGF-LIKE 20, CALCIUM-BINDING.
 FT DOMAIN 1318 1359 EGF-LIKE 21, CALCIUM-BINDING.
 FT DOMAIN 1360 1400 EGF-LIKE 22, CALCIUM-BINDING.
 FT DOMAIN 1401 1441 EGF-LIKE 23, CALCIUM-BINDING.
 FT DOMAIN 1442 1483 EGF-LIKE 24, CALCIUM-BINDING.
 FT DOMAIN 1484 1524 EGF-LIKE 25, CALCIUM-BINDING.
 FT DOMAIN 1525 1565 EGF-LIKE 26, CALCIUM-BINDING.
 FT REPEAT 1566 1642 TGFBBP 4.
 FT DOMAIN 1643 1684 EGF-LIKE 27, CALCIUM-BINDING.
 FT DOMAIN 1685 1726 EGF-LIKE 28, CALCIUM-BINDING.
 FT REPEAT 1727 1800 TGFBBP 5.
 FT DOMAIN 1801 1842 EGF-LIKE 29, CALCIUM-BINDING.
 FT DOMAIN 1843 1884 EGF-LIKE 30, CALCIUM-BINDING.
 FT DOMAIN 1885 1926 EGF-LIKE 31, CALCIUM-BINDING.
 FT DOMAIN 1927 1965 EGF-LIKE 32, CALCIUM-BINDING.
 FT DOMAIN 1966 2008 EGF-LIKE 33, CALCIUM-BINDING.
 FT DOMAIN 2009 2048 EGF-LIKE 34, CALCIUM-BINDING.
 FT DOMAIN 2049 2090 EGF-LIKE 35, CALCIUM-BINDING.
 FT REPEAT 2091 2163 TGFBBP 6.
 FT DOMAIN 2164 2205 EGF-LIKE 36, CALCIUM-BINDING.
 FT DOMAIN 2206 2245 EGF-LIKE 37, CALCIUM-BINDING.
 FT DOMAIN 2246 2286 EGF-LIKE 38, CALCIUM-BINDING.
 FT DOMAIN 2287 2330 EGF-LIKE 39, CALCIUM-BINDING.
 FT DOMAIN 2331 2372 EGF-LIKE 40, CALCIUM-BINDING.
 FT REPEAT 2373 2441 TGFBBP 7.
 FT DOMAIN 2442 2483 EGF-LIKE 41, CALCIUM-BINDING.
 FT DOMAIN 2484 2524 EGF-LIKE 42, CALCIUM-BINDING.
 FT DOMAIN 2525 2563 EGF-LIKE 43, CALCIUM-BINDING.
 FT DOMAIN 2564 2606 EGF-LIKE 44, CALCIUM-BINDING.
 FT DOMAIN 2607 2646 EGF-LIKE 45, CALCIUM-BINDING.
 FT DOMAIN 2647 2687 EGF-LIKE 46, CALCIUM-BINDING.
 FT DOMAIN 2688 2727 EGF-LIKE 47, CALCIUM-BINDING.
 FT DISULFID 115 124 BY SIMILARITY.
 FT DISULFID 119 130 BY SIMILARITY.
 FT DISULFID 132 141 BY SIMILARITY.
 FT DISULFID 149 159 BY SIMILARITY.
 FT DISULFID 153 164 BY SIMILARITY.
 FT DISULFID 166 175 BY SIMILARITY.
 FT DISULFID 180 190 BY SIMILARITY.
 FT DISULFID 184 196 BY SIMILARITY.
 FT DISULFID 198 207 BY SIMILARITY.
 FT DISULFID 280 292 BY SIMILARITY.
 FT DISULFID 287 301 BY SIMILARITY.
 FT DISULFID 303 316 BY SIMILARITY.
 FT DISULFID 322 334 BY SIMILARITY.
 FT DISULFID 329 343 BY SIMILARITY.
 FT DISULFID 345 358 BY SIMILARITY.
 FT DISULFID 491 503 BY SIMILARITY.
 FT DISULFID 498 512 BY SIMILARITY.
 FT DISULFID 514 526 BY SIMILARITY.
 FT DISULFID 532 542 BY SIMILARITY.
 FT DISULFID 537 551 BY SIMILARITY.
 FT DISULFID 553 566 BY SIMILARITY.
 FT DISULFID 572 584 BY SIMILARITY.
 FT DISULFID 579 593 BY SIMILARITY.
 FT DISULFID 595 608 BY SIMILARITY.
 FT DISULFID 614 625 BY SIMILARITY.
 FT DISULFID 620 634 BY SIMILARITY.
 FT DISULFID 636 649 BY SIMILARITY.
 FT DISULFID 655 666 BY SIMILARITY.
 FT DISULFID 661 675 BY SIMILARITY.
 FT DISULFID 677 690 BY SIMILARITY.
 FT DISULFID 765 777 BY SIMILARITY.
 FT DISULFID 772 786 BY SIMILARITY.
 FT DISULFID 788 801 BY SIMILARITY.
 FT DISULFID 807 819 BY SIMILARITY.
 FT DISULFID 814 828 BY SIMILARITY.
 FT DISULFID 830 843 BY SIMILARITY.
 FT DISULFID 849 859 BY SIMILARITY.
 FT DISULFID 854 868 BY SIMILARITY.
 FT DISULFID 870 883 BY SIMILARITY.
 FT DISULFID 952 964 BY SIMILARITY.
 FT DISULFID 959 973 BY SIMILARITY.
 FT DISULFID 975 988 BY SIMILARITY.
 FT DISULFID 1070 1082 BY SIMILARITY.
 FT DISULFID 1077 1091 BY SIMILARITY.
 FT DISULFID 1093 1106 BY SIMILARITY.
 FT DISULFID 1112 1124 BY SIMILARITY.
 FT DISULFID 1119 1133 BY SIMILARITY.
 FT DISULFID 1135 1149 BY SIMILARITY.
 FT DISULFID 1155 1167 BY SIMILARITY.
 FT DISULFID 1162 1176 BY SIMILARITY.
 FT DISULFID 1178 1191 BY SIMILARITY.
 FT DISULFID 1197 1209 BY SIMILARITY.
 FT DISULFID 1204 1218 BY SIMILARITY.
 FT DISULFID 1220 1233 BY SIMILARITY.
 FT DISULFID 1239 1250 BY SIMILARITY.
 FT DISULFID 1246 1259 BY SIMILARITY.
 FT DISULFID 1261 1274 BY SIMILARITY.
 FT DISULFID 1280 1292 BY SIMILARITY.
 FT DISULFID 1287 1301 BY SIMILARITY.
 FT DISULFID 1303 1316 BY SIMILARITY.
 FT DISULFID 1322 1334 BY SIMILARITY.
 FT DISULFID 1329 1343 BY SIMILARITY.
 FT DISULFID 1345 1358 BY SIMILARITY.
 FT DISULFID 1364 1377 BY SIMILARITY.
 FT DISULFID 1371 1386 BY SIMILARITY.
 FT DISULFID 1388 1399 BY SIMILARITY.
 FT DISULFID 1405 1418 BY SIMILARITY.
 FT DISULFID 1412 1427 BY SIMILARITY.
 FT DISULFID 1429 1440 BY SIMILARITY.
 FT DISULFID 1446 1458 BY SIMILARITY.
 FT DISULFID 1453 1467 BY SIMILARITY.
 FT DISULFID 1469 1482 BY SIMILARITY.
 FT DISULFID 1488 1499 BY SIMILARITY.
 FT DISULFID 1494 1508 BY SIMILARITY.
 FT DISULFID 1510 1523 BY SIMILARITY.
 FT DISULFID 1529 1540 BY SIMILARITY.
 FT DISULFID 1535 1549 BY SIMILARITY.
 FT DISULFID 1551 1564 BY SIMILARITY.
 FT DISULFID 1567 1579 BY SIMILARITY.
 FT DISULFID 1647 1659 BY SIMILARITY.
 FT DISULFID 1654 1668 BY SIMILARITY.
 FT DISULFID 1670 1683 BY SIMILARITY.
 FT DISULFID 1689 1701 BY SIMILARITY.

Query Match	13.4%	Score 410.5	DB 1	Length 2907
Best Local Similarity	28.4%	Pred. No. 1.2e-20		
Matches	95	Conservative 37	Mismatches 91	Indels 111
Gaps	13			
QY	38	PGVCHYGT-----KLACYGM-----RNSKGVCEATCEPG-CKFGECV---GPNKC 80		
Db	1074	PGMCTYKCRNTIGSFKCRNNGFALDMEERNCTDIDECRISPDLCSGIGCVNTPGSFEC 1133		
QY	81	RCFFGY-----TGKTCSQ-----D 94		
Db	1134	ECFEGYESGFMNMKNCMDIDECERNPILCRGGTCVNTGSGFQDCPLGHLSPSREDCVD 1193		
QY	95	VNECGMKPRPOH-RCVNTHGYSKCFCLSGHMLMPDAT-CVNSRTCAVIN--COYSCEDT 150		
Db	1194	INEGSLSDNLCNRNGKCNVMITGYQCSNPGYQATPDRGGCTDIDECMIMNGGCDTQCTNS 1253		
QY	151	EEGPOCLCPSSGLRLAPNGRDLIDCASKVTC-----PYNRCV-----192		
Db	1254	EGSYECSC-SEGVALMPDGRSCADIDCEENPDICDGGQCTNIPGEYRCLCYDGFMASMD 1312		
QY	193	-----NTFGSYCKCHIGFELQVIGRYVIGRIDINECTWDSHT 229		
Db	1313	MKTICIDVNECDLPNICMFGECENTKGSFICHQLGYSVK--XGTTGCTDVDECEICAHN 1370		
QY	230	CSHHANCFNTQGSFKCKCKQGYKNGLRCSAIPB 263		
Db	1371	CDMEASCLNVPGFSKCSREGWVGNGIKDLDLE 1404		
RESULT 2				
FBN2 HUMAN				
ID	FBN2 HUMAN	STANDARD;	PRT;	2911 AA.
AC	P35556;			
DT	01-JUN-1994 (Rel. 29, Created)			
DT	01-FEB-1996 (Rel. 33, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Fibrillin 2 precursor.			
GN	FBN2.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=94165150; PubMed=8120105;			
RA	Zhang H., Apfelroth S.D., Hu W., Davis E.C., Sanguinetti C.,			
RA	Bonadio J., Mecham R.P., Ramirez F.;			
RA	"Structure and expression of fibrillin-2, a novel microfibrillar			
RT	component preferentially located in elastic matrices.";			
RL	J. Cell Biol. 124:855-863(1994).			
RN	[2]			
RP	SEQUENCE OF 752-1505 FROM N.A.			
RX	MEDLINE=91304567; PubMed=1852206;			
RA	Lee B., Godfrey M., Vitale E., Hori H., Mattei M.-G., Sarfarazi M.,			
RA	Tsipouras P., Ramirez F., Hollister D.;			
RT	"Linkage of Marfan syndrome and a phenotypically related disorder to			
RL	two different fibrillin genes.";			
RL	Nature 352:330-334(1991).			
RN	[3]			
RP	VARIANTS CCA TYR-1252 AND SER-1433, AND VARIANT ILE-964.			
RX	MEDLINE=96083599; PubMed=7493032;			
RA	Putnam E.A., Zhang H., Ramirez F., Milewicz D.M.;			
RT	"Fibrillin-2 (FBN2) mutations result in the Marfan-like disorder,			
RL	congenital contractual arachnodactyly.";			
RL	Nat. Genet. 11:456-458(1995).			
RN	[4]			
RP	VARIANT CCA HIS-1114.			
RX	MEDLINE=98407789; PubMed=9737771;			
RA	Babcock D., Gaener C., Francke U., Maslen C.;			
RT	"A single mutation that results in an asp-to-his substitution and			
RL	partial exon skipping in a family with congenital contractual			
RT	arachnodactyly.";			

Hum. Genet. 103:22-28(1998).

[5]

RL VARIANTS CCA PHE-1141 AND TRP-1252.
RN MEDLINE=20259236; PubMed=10797416;
RX Belleh S., Zhou G., Wang M., Der Kaloustian V.M., Pagon R.A.,
RA Godfrey M.;
RT "Two novel fibrillin-2 mutations in congenital contractural
RT arachnodactyly.";
RL Am. J. Med. Genet. 92:7-12(2000).
CC -!- FUNCTION: Structural component of connective tissue microfibrils
CC that binds calcium. Fibrillin-2-containing microfibrils regulate
CC the early process of elastic fiber assembly.
CC -!- DISEASE: Defects in FBN2 are the cause of congenital contractural
CC arachnodactyly (CCA) [MIM:121050]; also known as Beals syndrome.
CC CCA is phenotypically similar to Marfan syndrome, but does not
CC effect the aorta and the eyes.
CC -!- SIMILARITY: Contains 47 EGF-like domains.
CC -!- SIMILARITY: Contains 7 TGF-beta binding protein (TGFBP) domains.
CC -!- DATABASE: NAME=Elastic Fiber Homepage; NOTE=fibrillin 2 page;
CC WWW="http://ef.wustl.edu/genes/FBN2.htm".

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; U03272; AAA18950.1; --
DR EMBL; X62009; -; NOT_ANNOTATED_CDS.
DR PTR; A54105; A54105.
DR HGSP; P35555; IEMN.
DR Genew; HGNC:3604; FBN2.
DR MIM; 121050; --
DR GT; GO:0005578; C:extracellular matrix; TAS.
DR GO; GO:0005201; F:extracellular matrix structural constituent; TAS.
DR GO; GO:0007345; F:embryogenesis and morphogenesis; TAS.
DR GO; GO:0007397; P:histogenesis and organogenesis; TAS.
DR InterPro; IPR000152; Asx_hydroxyl_s.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001436; EGF_II.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR002212; Fibrill-assoc.
DR Pfam; PF000008; EGF; 45.
DR Pfam; PF00683; TB; 9.
DR PRINTS; PR00010; EGF_BLOOD.
DR SMART; SM00179; EGF_CA; 43.
DR PROSITE; PS00010; ASX_HYDROXYL; 43.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS01186; EGF_2; 37.
DR PROSITE; PS50026; EGF_3; 45.
DR PROSITE; PS01187; EGF_CA; 42.
DR KW Extracellular matrix; Calcium-binding; Glycoprotein; EGF-like domain;
DR Repeat; Signal; Multigene family; Disease mutation; Polymorphism.
FT SIGNAL 1 28
FT CHAIN 29 2911
FT DOMAIN 111 142 FIBRILLIN 2.
FT DOMAIN 145 176 EGF-LIKE 1.
FT DOMAIN 176 207 EGF-LIKE 2.
FT DOMAIN 275 316 EGF-LIKE 3.
FT DOMAIN 317 358 EGF-LIKE 4.
FT REPEAT 359 425 EGF-LIKE 5, CALCIUM-BINDING.
FT DOMAIN 493 533 TGFBP 1.
FT DOMAIN 534 573 EGF-LIKE 6.
FT DOMAIN 574 615 EGF-LIKE 7, CALCIUM-BINDING.
FT DOMAIN 616 656 EGF-LIKE 8, CALCIUM-BINDING.
FT DOMAIN 657 697 EGF-LIKE 9, CALCIUM-BINDING.
FT REPEAT 698 766 TGFBP 2.
FT DOMAIN 767 808 EGF-LIKE 11, CALCIUM-BINDING.
FT DOMAIN 809 850 EGF-LIKE 12, CALCIUM-BINDING.
FT DOMAIN 851 890 EGF-LIKE 13, CALCIUM-BINDING.
FT DOMAIN 954 995 EGF-LIKE 14, CALCIUM-BINDING.


```

-!- FUNCTION: Structural component of connective tissue microfibrils
CC that binds calcium. Fibrillin-1-containing microfibrils provide
CC long-term force bearing structural support.
CC
-!- PWM: Forms intermolecular disulfide bonds either with other
CC fibrillin-1 molecules or with other components of the
CC microfibrils.
CC
-!- SIMILARITY: Contains 47 EGF-like domains.
CC
-!- SIMILARITY: Contains 7 TGF-beta binding protein (TGFBP) domains.
CC
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
EMBL: L28748; AAA74122.1; -.
DR
PIR: A55567; A55567.
DR
HSP; P35555; IAPU.
DR
InterPro: IPR000152; Asx hydroxyl_S.
DR
InterPro: IPR001881; EGF_Ca.
DR
InterPro: IPR006209; EGF-like.
DR
InterPro: IPR002212; Fibril-assoc.
DR
Pfam: PF00008; EGF; 46.
DR
Pfam: PF00693; TB; 9.
DR
SMART: SM00179; EGF CA; 42.
DR
PROSITE; PS00010; ASX HYDROXYL; 43.
DR
PROSITE; PS00022; EGF_1; 2.
DR
PROSITE; PS01186; EGF_2; 38.
DR
PROSITE; PS00026; EGF_3; 45.
DR
PROSITE; PS01187; EGF CA; 43.
DR
Extracellular matrix; Calcium-binding; Glycoprotein; EGF-like domain;
KW Repeat; Signal; Multigene family.
CC
SIGNAL 1 27 POTENTIAL.
FT CHAIN 28 2871 FIBRILLIN 1.
FT DOMAIN 81 112 EGF-LIKE 1.
FT DOMAIN 115 146 EGF-LIKE 2.
FT DOMAIN 147 178 EGF-LIKE 3.
FT DOMAIN 246 287 EGF-LIKE 4.
FT DOMAIN 288 329 EGF-LIKE 5, CALCIUM-BINDING.
FT REPEAT 330 390 EGF-LIKE 5, CALCIUM-BINDING.
FT DOMAIN 392 446 TGFBP 1.
FT DOMAIN 449 489 PRO-RICH.
FT DOMAIN 490 529 EGF-LIKE 6.
FT DOMAIN 530 571 EGF-LIKE 7.
FT DOMAIN 572 612 EGF-LIKE 8.
FT DOMAIN 613 653 EGF-LIKE 9.
FT REPEAT 654 722 EGF-LIKE 10, CALCIUM-BINDING.
FT DOMAIN 723 764 EGF-LIKE 11, CALCIUM-BINDING.
FT DOMAIN 765 806 EGF-LIKE 12, CALCIUM-BINDING.
FT DOMAIN 807 846 EGF-LIKE 13, CALCIUM-BINDING.
FT DOMAIN 910 951 EGF-LIKE 14, CALCIUM-BINDING.
FT REPEAT 952 1027 TGFBP 3.
FT DOMAIN 1028 1069 EGF-LIKE 15, CALCIUM-BINDING.
FT DOMAIN 1070 1112 EGF-LIKE 16, CALCIUM-BINDING.
FT DOMAIN 1113 1154 EGF-LIKE 17, CALCIUM-BINDING.
FT DOMAIN 1155 1196 EGF-LIKE 18, CALCIUM-BINDING.
FT DOMAIN 1197 1237 EGF-LIKE 19, CALCIUM-BINDING.
FT DOMAIN 1238 1279 EGF-LIKE 20, CALCIUM-BINDING.
FT DOMAIN 1280 1321 EGF-LIKE 21, CALCIUM-BINDING.
FT DOMAIN 1322 1362 EGF-LIKE 22, CALCIUM-BINDING.
FT DOMAIN 1363 1403 EGF-LIKE 23, CALCIUM-BINDING.
FT DOMAIN 1404 1445 EGF-LIKE 24, CALCIUM-BINDING.
FT DOMAIN 1446 1486 EGF-LIKE 25, CALCIUM-BINDING.
FT DOMAIN 1487 1527 EGF-LIKE 26, CALCIUM-BINDING.
FT REPEAT 1528 1605 TGFBP 4.
FT DOMAIN 1606 1647 EGF-LIKE 27, CALCIUM-BINDING.
FT DOMAIN 1648 1688 EGF-LIKE 28, CALCIUM-BINDING.
FT REPEAT 1689 1765 TGFBP 5.
FT DOMAIN 1766 1807 EGF-LIKE 29, CALCIUM-BINDING.
FT DOMAIN 1808 1848 EGF-LIKE 30, CALCIUM-BINDING.
FT DOMAIN 1849 1890 EGF-LIKE 31, CALCIUM-BINDING.

```

```

FT DISULFID 1208 1221 BY SIMILARITY.
FT DISULFID 1223 1236 BY SIMILARITY.
FT DISULFID 1242 1254 BY SIMILARITY.
FT DISULFID 1249 1263 BY SIMILARITY.
FT DISULFID 1265 1278 BY SIMILARITY.
FT DISULFID 1284 1296 BY SIMILARITY.
FT DISULFID 1291 1305 BY SIMILARITY.
FT DISULFID 1307 1320 BY SIMILARITY.
FT DISULFID 1326 1339 BY SIMILARITY.
FT DISULFID 1333 1348 BY SIMILARITY.
FT DISULFID 1350 1361 BY SIMILARITY.
FT DISULFID 1367 1380 BY SIMILARITY.
FT DISULFID 1374 1389 BY SIMILARITY.
FT DISULFID 1391 1402 BY SIMILARITY.
FT DISULFID 1408 1420 BY SIMILARITY.
FT DISULFID 1415 1429 BY SIMILARITY.
FT DISULFID 1431 1444 BY SIMILARITY.
FT DISULFID 1450 1461 BY SIMILARITY.
FT DISULFID 1456 1470 BY SIMILARITY.
FT DISULFID 1472 1485 BY SIMILARITY.
FT DISULFID 1491 1502 BY SIMILARITY.
FT DISULFID 1497 1511 BY SIMILARITY.
FT DISULFID 1513 1526 BY SIMILARITY.
FT DISULFID 1610 1622 BY SIMILARITY.
FT DISULFID 1617 1631 BY SIMILARITY.
FT DISULFID 1633 1646 BY SIMILARITY.

Query Match 12.7%; Score 387.5; DB 1; Length 2871;
Best Local Similarity 30.6%; Pred. No. 5e-19;
Matches 89; Conservative 34; Mismatches 77; Gaps 10;

QY 49 CCYGNRRNSKGVCRATCPG-----CKFGCV---GPNKRC 82
Db 1081 CGRGQVNTPGDFECKDEGVESGPMNMKNCDIDECQDPLCRGGVCLNTEGSRCEC 1140

QY 83 FPGY---TGKTCSDVNECGMKRCPQH-RCVNTHSGYKFCLSHMLMPDAT-CVNSRT 137
Db 1141 PFHQLANISACIDINECELSAHLCPHGRCVNLIGYQACNPGYHSTPRLFCVIDE 1200

QY 138 CAMIN--CQYSCETEPQCLCPSSGLRLAPNGRDCLDIDECASGVIC-----PY 187
Db 1201 CSIMNGCETCTNSESGYECSC-QPGFALPDQQRCTDIDECEDNPNICDGGQCTNIPG 1259

QY 188 NRCV-----NTGSSYKCHIGFELYISGR 214
Db 1260 EYRCLVDGFWASEDMKTCVDVNECDLNPICLSGTCENTKGSFICHCDMGYSK--KGK 1317

QY 215 YDCIDINECTMDSTCSHANCFTQGSFKCKQGYKNGLRCSAIPENS 265
Db 1318 TGCTDINECEGAHNCDRHAVCTNTAGSFKCSFGWIGDKICTDLDECS 1368

RESULT 4
ID FBN1 MOUSE STANDARD; PRT; 2871 AA.
AC Q61554; Q60826;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Fibrillin 1 precursor.
GN FBN1 OR FBN-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95130561; PubMed=7829516;
RA Yin W., Germiller J., Sanguinetti C., Smiley E., Pangilinan T.,
RA Pereira L., Ramirez F., Bonadio J.;
RT "Primary structure and developmental expression of Fbn-1, the mouse
RT fibrillin gene."
RL J. Biol. Chem. 270:1798-1806(1995).

```

```

RN RC SEQUENCE FROM N.A.
RA STRAIN=CD-1; TISSUE=Kidney;
RA Ota K, Kumar A, Wada J, Liu Z, Kanwar Y.S.;
RA Submitted (APR-1995) to the EMBL/Genbank/DBJ databases.
CC -!- FUNCTION: Structural component of connective tissue microfibrils
CC that binds calcium. Fibrillin-1-containing microfibrils provide
CC long-term force bearing structural support.
CC -!- PTM: Forms intermolecular disulfide bonds either with other
CC fibrillin-1 molecules or with other components of the microfibrils
CC (By similarity).
CC -!- SIMILARITY: Contains 47 EGF-like domains.
CC -!- SIMILARITY: Contains 7 TGF-beta binding protein (TGFBP) domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L29454; AAA56840.1; --
DR EMBL; U22493; AAA64217.1; --
DR PIR; A55624; A55624.
DR HSP; P35555; IAPJ.
DR MGD; MGI:95489; Fbn1.
DR InterPro; IPR000152; Asx_hydroxyl_S.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR006209; EGF_Like.
DR InterPro; IPR002212; Fibril-associ.
DR Pfam; PF00008; EGF; 45.
DR Pfam; PF00683; TB; 9.
DR SMART; SM00179; EGF_CA; 42.
DR PROSITE; PS00010; ASX_HYDROXYL; 43.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS01186; EGF_2; 38.
DR PROSITE; PS00026; EGF_3; 45.
DR PROSITE; PS01187; EGF_CA; 43.
DR Extracellular matrix; Calcium-binding; Glycoprotein; EGF-like domain;
KW Repeat; Signal; Multigene family.
FT SIGNAL 1 27 POTENTIAL.
FT CHAIN 28 2871 FIBRILLIN 1.
FT DOMAIN 81 112 EGF-LIKE 1.
FT DOMAIN 115 146 EGF-LIKE 2.
FT DOMAIN 147 178 EGF-LIKE 3.
FT DOMAIN 246 287 EGF-LIKE 4.
FT DOMAIN 288 329 EGF-LIKE 5.
FT DOMAIN 330 401 TGFBP 1.
FT DOMAIN 402 446 PRO-RICH.
FT DOMAIN 449 489 EGF-LIKE 6.
FT DOMAIN 490 529 EGF-LIKE 7.
FT DOMAIN 530 571 EGF-LIKE 8.
FT DOMAIN 572 612 EGF-LIKE 9.
FT DOMAIN 613 653 EGF-LIKE 10.
FT DOMAIN 656 721 TGFBP 2.
FT DOMAIN 723 764 EGF-LIKE 11.
FT DOMAIN 765 806 EGF-LIKE 12.
FT DOMAIN 807 846 EGF-LIKE 13.
FT DOMAIN 910 951 EGF-LIKE 14.
FT DOMAIN 952 1018 TGFBP 3.
FT DOMAIN 1028 1069 EGF-LIKE 15.
FT DOMAIN 1070 1112 EGF-LIKE 16.
FT DOMAIN 1113 1154 EGF-LIKE 17.
FT DOMAIN 1155 1196 EGF-LIKE 18.
FT DOMAIN 1197 1237 EGF-LIKE 19.
FT DOMAIN 1238 1279 EGF-LIKE 20.
FT DOMAIN 1280 1321 EGF-LIKE 21.
FT DOMAIN 1322 1362 EGF-LIKE 22.
FT DOMAIN 1363 1403 EGF-LIKE 23.
FT DOMAIN 1404 1445 EGF-LIKE 24.
FT DOMAIN 1446 1486 EGF-LIKE 25.
FT DOMAIN 1487 1527 EGF-LIKE 26.

```

```
FT DOMAIN 1528 1599 TGRBP 4. EGF-LIKE 27, CALCIUM-BINDING.
FT DOMAIN 1606 1647 EGF-LIKE 28, CALCIUM-BINDING.
FT DOMAIN 1648 1688 TGRBP 5. EGF-LIKE 29, CALCIUM-BINDING.
FT DOMAIN 1689 1758 EGF-LIKE 30, CALCIUM-BINDING.
FT DOMAIN 1766 1807 EGF-LIKE 31, CALCIUM-BINDING.
FT DOMAIN 1808 1848 EGF-LIKE 32, CALCIUM-BINDING.
FT DOMAIN 1849 1890 EGF-LIKE 33, CALCIUM-BINDING.
FT DOMAIN 1891 1929 EGF-LIKE 34, CALCIUM-BINDING.
FT DOMAIN 1930 1972 EGF-LIKE 35, CALCIUM-BINDING.
FT DOMAIN 1973 2012 TGRBP 6. EGF-LIKE 36, CALCIUM-BINDING.
FT DOMAIN 2013 2054 EGF-LIKE 37, CALCIUM-BINDING.
FT DOMAIN 2055 2121 EGF-LIKE 38, CALCIUM-BINDING.
FT DOMAIN 2127 2165 EGF-LIKE 39, CALCIUM-BINDING.
FT DOMAIN 2166 2205 EGF-LIKE 40, CALCIUM-BINDING.
FT DOMAIN 2206 2246 TGRBP 7. EGF-LIKE 41, CALCIUM-BINDING.
FT DOMAIN 2247 2290 EGF-LIKE 42, CALCIUM-BINDING.
FT DOMAIN 2291 2332 EGF-LIKE 43, CALCIUM-BINDING.
FT DOMAIN 2333 2400 EGF-LIKE 44, CALCIUM-BINDING.
FT DOMAIN 2402 2443 EGF-LIKE 45, CALCIUM-BINDING.
FT DOMAIN 2444 2484 EGF-LIKE 46, CALCIUM-BINDING.
FT DOMAIN 2485 2523 EGF-LIKE 47, CALCIUM-BINDING.
FT DOMAIN 2524 2566 EGF-LIKE 48, CALCIUM-BINDING.
FT DOMAIN 2567 2606 EGF-LIKE 49, CALCIUM-BINDING.
FT DOMAIN 2607 2647 EGF-LIKE 50, CALCIUM-BINDING.
FT DOMAIN 2648 2687 EGF-LIKE 51, CALCIUM-BINDING.
FT DISULFID 85 94 BY SIMILARITY.
FT DISULFID 89 100 BY SIMILARITY.
FT DISULFID 102 111 BY SIMILARITY.
FT DISULFID 119 129 BY SIMILARITY.
FT DISULFID 123 134 BY SIMILARITY.
FT DISULFID 136 145 BY SIMILARITY.
FT DISULFID 150 166 BY SIMILARITY.
FT DISULFID 154 166 BY SIMILARITY.
FT DISULFID 168 177 BY SIMILARITY.
FT DISULFID 250 262 BY SIMILARITY.
FT DISULFID 257 271 BY SIMILARITY.
FT DISULFID 273 286 BY SIMILARITY.
FT DISULFID 292 304 BY SIMILARITY.
FT DISULFID 299 313 BY SIMILARITY.
FT DISULFID 315 328 BY SIMILARITY.
FT DISULFID 453 465 BY SIMILARITY.
FT DISULFID 460 474 BY SIMILARITY.
FT DISULFID 476 488 BY SIMILARITY.
FT DISULFID 494 504 BY SIMILARITY.
FT DISULFID 499 513 BY SIMILARITY.
FT DISULFID 515 528 BY SIMILARITY.
FT DISULFID 534 546 BY SIMILARITY.
FT DISULFID 541 555 BY SIMILARITY.
FT DISULFID 557 570 BY SIMILARITY.
FT DISULFID 576 587 BY SIMILARITY.
FT DISULFID 582 596 BY SIMILARITY.
FT DISULFID 598 611 BY SIMILARITY.
FT DISULFID 617 628 BY SIMILARITY.
FT DISULFID 623 637 BY SIMILARITY.
FT DISULFID 639 652 BY SIMILARITY.
FT DISULFID 727 739 BY SIMILARITY.
FT DISULFID 734 748 BY SIMILARITY.
FT DISULFID 750 763 BY SIMILARITY.
FT DISULFID 769 781 BY SIMILARITY.
FT DISULFID 776 790 BY SIMILARITY.
FT DISULFID 792 805 BY SIMILARITY.
FT DISULFID 811 821 BY SIMILARITY.
FT DISULFID 816 830 BY SIMILARITY.
FT DISULFID 832 845 BY SIMILARITY.
FT DISULFID 914 926 BY SIMILARITY.
FT DISULFID 921 935 BY SIMILARITY.
FT DISULFID 937 950 BY SIMILARITY.
FT DISULFID 1032 1044 BY SIMILARITY.
FT DISULFID 1039 1053 BY SIMILARITY.
FT DISULFID 1055 1068 BY SIMILARITY.
FT DISULFID 1074 1086 BY SIMILARITY.
FT DISULFID 1081 1095 BY SIMILARITY.
FT DISULFID 1097 1111 BY SIMILARITY.
```

```
FT DISULFID 1117 1129 BY SIMILARITY.
FT DISULFID 1124 1138 BY SIMILARITY.
FT DISULFID 1140 1153 BY SIMILARITY.
FT DISULFID 1159 1171 BY SIMILARITY.
FT DISULFID 1166 1180 BY SIMILARITY.
FT DISULFID 1182 1195 BY SIMILARITY.
FT DISULFID 1201 1212 BY SIMILARITY.
FT DISULFID 1208 1221 BY SIMILARITY.
FT DISULFID 1223 1236 BY SIMILARITY.
FT DISULFID 1242 1254 BY SIMILARITY.
FT DISULFID 1249 1263 BY SIMILARITY.
FT DISULFID 1265 1278 BY SIMILARITY.
FT DISULFID 1284 1296 BY SIMILARITY.
FT DISULFID 1291 1305 BY SIMILARITY.
FT DISULFID 1307 1320 BY SIMILARITY.
FT DISULFID 1326 1339 BY SIMILARITY.
FT DISULFID 1333 1348 BY SIMILARITY.
FT DISULFID 1350 1361 BY SIMILARITY.
FT DISULFID 1367 1380 BY SIMILARITY.
FT DISULFID 1374 1389 BY SIMILARITY.
FT DISULFID 1391 1402 BY SIMILARITY.
FT DISULFID 1408 1420 BY SIMILARITY.
FT DISULFID 1415 1429 BY SIMILARITY.
FT DISULFID 1431 1444 BY SIMILARITY.
FT DISULFID 1450 1461 BY SIMILARITY.
FT DISULFID 1456 1470 BY SIMILARITY.
FT DISULFID 1472 1485 BY SIMILARITY.
FT DISULFID 1491 1502 BY SIMILARITY.
FT DISULFID 1497 1511 BY SIMILARITY.
FT DISULFID 1513 1526 BY SIMILARITY.
FT DISULFID 1610 1622 BY SIMILARITY.
FT DISULFID 1617 1631 BY SIMILARITY.
FT DISULFID 1633 1646 BY SIMILARITY.
FT DISULFID 1652 1663 BY SIMILARITY.
FT DISULFID 1658 1672 BY SIMILARITY.

Query Match 12.6%; Score 385.5; DB 1; Length 2871;
Best Local Similarity 32.2%; Pred. No. 6.9e-19;
Matches 86; Conservative 34; Mismatches 82; Indels 65; Gaps 11;

QY 39 GVCHYGTKLACCYGVRRNSKGVCEATCPG-----NTEGYRCECPGQLSNISACIDINECELSANLCPHGRCVN 1173
Db 1127 GICH-----NTEGYRCECPGQLSNISACIDINECELSANLCPHGRCVN 1173

QY 76 --GPNKRCRCPGY---TGKTCSDVNECMKPRPCQHRVCVNTGHSYKCFCLSGHMLMPDA 130
Db 1174 LIGKYQCACNPGYHPTDRLFCVDIDECISNNGGCTETCTNSDGSYECSCQPGFALMPD- 1232

QY 131 TCVNSRTAMINCQYSCEDTE---BGPO-----CLCPSSGLRLAPNGRDCLDIDEC 178
Db 1233 ----QRSCTDID---QCEDNPNICDGGQCTNIPGEYRCLC-YDGFMASEDMMKTCYDVNRC 1284

QY 179 ASGKVICPNRRVCNTFGSYCKHGFELQVYISGRYDCIDINECTMDSHSTCSHHANCFN 238
Db 1285 DLNPNIC-LSGTCENTKSFICHCDMGYSKG--KGKTGCTDINECEIHAHNCGRAVCTN 1341

QY 239 TQGSFKCKKQYKGNGLRCSAIPENS 265
Db 1342 TAGSFKCSGPGWIGDKICTDLDECS 1368

RESULT 5
FBNI_PIG ID FBNI_PIG STANDARD; PRT; 2871 AA.
AC Q9TV36; DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Fibrillin 1 precursor.
GN FBNI.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
```


OX NCBI_TaxID=9823;
 RP [1]
 RN SEQUENCE FROM N.A.
 RC TISSUE=Lung;
 RX MEDLINE=99156858; PubMed=10036187;
 RA Biery N.J., Eldadah Z.A., Moore C.S., Stetten G., Spencer F.,
 RA Dietz H.C.;
 RT "Revised genomic organization of FBN1 and significance for regulated
 RL gene expression.";
 RL Genomics 56:70-77(1999).
 CC -!- FUNCTION: Structural component of connective tissue microfibrils
 CC that binds calcium. Fibrillin-1-containing microfibrils provide
 CC long-term force bearing structural support.
 CC -!- PTM: Forms intermolecular disulfide bonds either with other
 CC fibrillin-1 molecules or with other components of the
 CC microfibrils.
 CC -!- SIMILARITY: Contains 47 EGF-like domains.
 CC -!- SIMILARITY: Contains 7 TGF-beta binding protein (TGFBP) domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR HSSP; AF073800; AAD50328.1; -;
 DR HSSP; P35555; IAPJ.
 DR InterPro; IPR00152; Asx hydroxyl_s.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR006209; EGF_Like.
 DR InterPro; IPR002212; Fibril-assoc.
 DR Pfam; PF00008; EGF; 44.
 DR Pfam; PF00683; TB; 9.
 DR SMART; SM00179; EGF_CA; 40.
 DR PROSITE; PS00010; ASX_HYDROXYL; 41.
 DR PROSITE; PS00022; EGF_1; 2.
 DR PROSITE; PS01186; EGF_2; 36.
 DR PROSITE; PS00026; EGF_3; 43.
 DR PROSITE; PS01187; EGF_CA; 41.
 DR Extracellular matrix; Calcium-binding; Glycoprotein; EGF-like domain;
 KW Repeat; Signal; Multigene family.
 FT SIGNAL 1 27 POTENTIAL.
 FT CHAIN 28 2871 FIBRILLIN 1.
 FT DOMAIN 81 112 EGF-LIKE 1.
 FT DOMAIN 115 146 EGF-LIKE 2.
 FT DOMAIN 147 178 EGF-LIKE 3.
 FT DOMAIN 246 287 EGF-LIKE 4.
 FT DOMAIN 288 329 EGF-LIKE 5.
 FT REPEAT 330 390 TGFBP 1.
 FT DOMAIN 392 446 PRO-RICH.
 FT DOMAIN 449 489 EGF-LIKE 6.
 FT DOMAIN 490 529 EGF-LIKE 7.
 FT DOMAIN 530 571 EGF-LIKE 8.
 FT DOMAIN 572 612 EGF-LIKE 9.
 FT DOMAIN 613 653 EGF-LIKE 10.
 FT REPEAT 654 722 TGFBP 2.
 FT DOMAIN 723 764 EGF-LIKE 11.
 FT DOMAIN 765 806 EGF-LIKE 12.
 FT DOMAIN 807 846 EGF-LIKE 13.
 FT DOMAIN 910 951 EGF-LIKE 14.
 FT REPEAT 952 1027 TGFBP 3.
 FT DOMAIN 1028 1069 EGF-LIKE 15.
 FT DOMAIN 1070 1112 EGF-LIKE 16.
 FT DOMAIN 1113 1154 EGF-LIKE 17.
 FT DOMAIN 1155 1196 EGF-LIKE 18.
 FT DOMAIN 1197 1237 EGF-LIKE 19.
 FT DOMAIN 1238 1279 EGF-LIKE 20.
 FT DOMAIN 1280 1321 EGF-LIKE 21.
 FT DOMAIN 1322 1362 EGF-LIKE 22.
 FT DOMAIN 1363 1403 EGF-LIKE 23.
 FT DOMAIN 1404 1445 EGF-LIKE 24.
 FT DOMAIN 1486 1486 EGF-LIKE 25.
 FT DOMAIN 1487 1527 EGF-LIKE 26.
 FT REPEAT 1528 1605 TGFBP 4.
 FT DOMAIN 1606 1647 EGF-LIKE 27.
 FT DOMAIN 1648 1688 EGF-LIKE 28.
 FT REPEAT 1689 1765 TGFBP 5.
 FT DOMAIN 1766 1807 EGF-LIKE 29.
 FT DOMAIN 1808 1848 EGF-LIKE 30.
 FT DOMAIN 1849 1890 EGF-LIKE 31.
 FT DOMAIN 1891 1929 EGF-LIKE 32.
 FT DOMAIN 1930 1972 EGF-LIKE 33.
 FT DOMAIN 1973 2012 EGF-LIKE 34.
 FT DOMAIN 2013 2054 EGF-LIKE 35.
 FT REPEAT 2055 2126 TGFBP 6.
 FT DOMAIN 2127 2165 EGF-LIKE 36.
 FT DOMAIN 2166 2205 EGF-LIKE 37.
 FT DOMAIN 2206 2246 EGF-LIKE 38.
 FT DOMAIN 2247 2290 EGF-LIKE 39.
 FT DOMAIN 2291 2332 EGF-LIKE 40.
 FT REPEAT 2333 2401 TGFBP 7.
 FT DOMAIN 2402 2443 EGF-LIKE 41.
 FT DOMAIN 2444 2484 EGF-LIKE 42.
 FT DOMAIN 2485 2524 EGF-LIKE 43.
 FT DOMAIN 2524 2566 EGF-LIKE 44.
 FT DOMAIN 2567 2606 EGF-LIKE 45.
 FT DOMAIN 2607 2647 EGF-LIKE 46.
 FT DOMAIN 2648 2687 EGF-LIKE 47.
 FT DISULFID 85 94 BY SIMILARITY.
 FT DISULFID 89 100 BY SIMILARITY.
 FT DISULFID 102 111 BY SIMILARITY.
 FT DISULFID 119 129 BY SIMILARITY.
 FT DISULFID 123 134 BY SIMILARITY.
 FT DISULFID 136 145 BY SIMILARITY.
 FT DISULFID 150 160 BY SIMILARITY.
 FT DISULFID 154 166 BY SIMILARITY.
 FT DISULFID 168 177 BY SIMILARITY.
 FT DISULFID 250 262 BY SIMILARITY.
 FT DISULFID 257 271 BY SIMILARITY.
 FT DISULFID 273 286 BY SIMILARITY.
 FT DISULFID 292 304 BY SIMILARITY.
 FT DISULFID 299 313 BY SIMILARITY.
 FT DISULFID 315 328 BY SIMILARITY.
 FT DISULFID 453 465 BY SIMILARITY.
 FT DISULFID 460 474 BY SIMILARITY.
 FT DISULFID 476 488 BY SIMILARITY.
 FT DISULFID 494 504 BY SIMILARITY.
 FT DISULFID 499 513 BY SIMILARITY.
 FT DISULFID 515 528 BY SIMILARITY.
 FT DISULFID 534 546 BY SIMILARITY.
 FT DISULFID 541 555 BY SIMILARITY.
 FT DISULFID 557 570 BY SIMILARITY.
 FT DISULFID 576 587 BY SIMILARITY.
 FT DISULFID 582 596 BY SIMILARITY.
 FT DISULFID 598 611 BY SIMILARITY.
 FT DISULFID 617 628 BY SIMILARITY.
 FT DISULFID 623 637 BY SIMILARITY.
 FT DISULFID 639 652 BY SIMILARITY.
 FT DISULFID 727 739 BY SIMILARITY.
 FT DISULFID 734 748 BY SIMILARITY.
 FT DISULFID 750 763 BY SIMILARITY.
 FT DISULFID 769 781 BY SIMILARITY.
 FT DISULFID 776 790 BY SIMILARITY.
 FT DISULFID 792 805 BY SIMILARITY.
 FT DISULFID 811 821 BY SIMILARITY.
 FT DISULFID 816 830 BY SIMILARITY.
 FT DISULFID 832 845 BY SIMILARITY.
 FT DISULFID 914 926 BY SIMILARITY.
 FT DISULFID 921 935 BY SIMILARITY.
 FT DISULFID 937 950 BY SIMILARITY.
 FT DISULFID 1032 1044 BY SIMILARITY.
 FT DISULFID 1039 1053 BY SIMILARITY.
 FT DISULFID 1055 1068 BY SIMILARITY.
 FT DISULFID 1074 1086 BY SIMILARITY.

factor-like domains: implications for the Marfan syndrome and other genetic disorders.";
 RL Cell 85:597-605(1996).
 RN [9]
 RP REVIEW ON MFS VARIANTS.
 RP MEDLINE=96174615; PubMed=8594563;
 RA Colloid G., Beroud C., Soussi T., Junien C., Boileau C.;
 RT "Software and database for the analysis of mutations in the human
 RT FBN1 gene.";
 RL Nucleic Acids Res. 24:137-141(1996).
 RN [10]
 RP REVIEW ON MFS VARIANTS.
 RP MEDLINE=97169383; PubMed=9016526;
 RA Colloid-Beroud G., Beroud C., Ades L., Black C., Boxer D.J.,
 RA Godfrey M., Hayward C., Karttunen L., Milewicz D., Peltonen L.,
 RA Richards R.I., Wang W., Junien C., Boileau C.;
 RT "Marfan database (second edition): software and database for the
 RT analysis of mutations in the human FBN1 gene.";
 RL Nucleic Acids Res. 25:147-150(1997).
 RN [11]
 RP REVIEW ON VARIANTS.
 RP MEDLINE=98062175; PubMed=9401003;
 RA Hayward C., Brock D.J.H.;
 RT "Fibrillin-1 mutations in Marfan syndrome and other type-1
 RT fibrillinopathies.";
 RL Hum. Mutat. 10:415-423(1997).
 RN [12]
 RP VARIANT MFS PRO-1137.
 RP MEDLINE=91304569; PubMed=1852208;
 RA Dietz H.C., Cutting G.R., Pyeritz R.E., Maslen C.L., Sakai L.Y.,
 RA Corson G.M., Puffenberger E.G., Hamosh A., Nanthakumar E.J.,
 RA Curristin S.M., Stetten G., Meyers D.A., Francomano C.A.;
 RT "Marfan syndrome caused by a recurrent de novo missense mutation in
 RT the fibrillin gene.";
 RL Nature 352:337-339(1991).
 RN [13]
 RP VARIANTS MFS SER-1249; ARG-1663; SER-2221 AND SER-2307.
 RP MEDLINE=93250834; PubMed=1301946;
 RA Dietz H.C., Saraiva J.M., Pyeritz R.E., Cutting G.R., Francomano C.A.;
 RT "Clustering of fibrillin (FBN1) missense mutations in Marfan syndrome
 RT patients at cysteine residues in EGF-like domains.";
 RL Hum. Mutat. 1:366-374(1992).
 RN [14]
 RP VARIANT MFS SER-2307.
 RP MEDLINE=92235290; PubMed=1569206;
 RA Dietz H.C., Pyeritz R.E., Puffenberger E.G., Kendzior R.J. Jr.,
 RA Corson G.M., Maslen C.L., Sakai L.Y., Francomano C.A., Cutting G.R.;
 RT "Marfan phenotype variability in a family segregating a missense
 RT mutation in the epidermal growth factor-like motif of the fibrillin
 RT gene.";
 RL J. Clin. Invest. 89:1674-1680(1992).
 RN [15]
 RP VARIANTS MFS ILE-548 AND ALA-723.
 RP MEDLINE=94010946; PubMed=8406497;
 RA Dietz H.C., McIntosh I., Sakai L.Y., Corson G.M., Chalberg S.C.,
 RA Pyeritz R.E., Francomano C.A.;
 RT "Four novel FBN1 mutations: significance for mutant transcript level
 RT and EGF-like domain calcium binding in the pathogenesis of Marfan
 RT syndrome.";
 RL Genomics 17:468-475(1993).
 RN [16]
 RP VARIANT MFS SER-2144.
 RP MEDLINE=93278402; PubMed=8504310;
 RA Hewett D.R., Lynch J.R., Smith R., Sykes B.C.;
 RT "A novel fibrillin mutation in the Marfan syndrome which could
 RT disrupt calcium binding of the epidermal growth factor-like module.";
 RL Hum. Mol. Genet. 2:475-477(1993).
 RN [17]
 RP VARIANTS MFS ARG-862; TYR-1117; PRO-1137 AND PHE-1589, AND VARIANT
 RP ALA-1148.
 RP MEDLINE=94108431; PubMed=8281141;
 RA Tynan K., Comeau K., Pearson M., Wilgenbus P., Levitt D., Gasner C.,
 RA Berg M.A., Miller D.C., Francke U.;

"Mutation screening of complete fibrillin-1 coding sequence: report
 of five new mutations, including two in 8-cysteine domains.";
 Hum. Mol. Genet. 2:1813-1821(1993).
 RN [18]
 RP VARIANTS MFS GLY-217 AND ARG-2627.
 RP MEDLINE=95067970; PubMed=7977366;
 RA Karttunen L., Raghunath M., Loenneqvist L., Peltonen L.;
 RT "A compound-heterozygous Marfan patient: two defective fibrillin
 RT alleles result in a lethal phenotype.";
 RL Am. J. Hum. Genet. 55:1083-1091(1994).
 RN [19]
 RP VARIANT EL LYS-2447.
 RP MEDLINE=94245249; PubMed=8198302;
 RA Lonngvist L., Child A., Kainulainen K., Davidson R., Puhakka L.,
 RA Peltonen L.;
 RT "A novel mutation of the fibrillin gene causing ectopia lentis.";
 RL Genomics 19:573-576(1994).
 RN [20]
 RP VARIANT MFS CYS-627.
 RP MEDLINE=94272487; PubMed=8004112;
 RA Hayward C., Rae A.L., Porteous M.E.M., Logie L.J., Brock L.J.;
 RT "Two novel mutations and a neutral polymorphism in EGF-like domains
 RT of the fibrillin gene (FBN1): SSCP screening of exons 15-21 in Marfan
 RT syndrome patients.";
 RL Hum. Mol. Genet. 3:373-375(1994).
 RN [21]
 RP VARIANT MFS CYS-122.
 RP MEDLINE=94314977; PubMed=8040326;
 RA Stahl-Hallengren C., Ukkonen T., Kainulainen K., Kristofersson U.,
 RA Saxne T., Torqvist K., Peltonen L.;
 RT "An extra cysteine in one of the non-calcium-binding epidermal growth
 RT factor-like motifs of the FBN1 polypeptide is connected to a novel
 RT variant of Marfan syndrome.";
 RL J. Clin. Invest. 94:709-713(1994).
 RN [22]
 RP VARIANT MFS TYR-1223.
 RP MEDLINE=94351682; PubMed=8071963;
 RA Hewett D.R., Lynch J.R., Child A., Sykes B.C.;
 RT "A new missense mutation of fibrillin in a patient with Marfan
 RT syndrome.";
 RL J. Med. Genet. 31:338-339(1994).
 RN [23]
 RP VARIANT MFS HIS-1170.
 RP MEDLINE=95174777; PubMed=7870075;
 RA Hayward C., Porteous M.E.M., Brock D.J.H.;
 RT "A novel mutation in the fibrillin gene (FBN1) in familial
 RT arachnodactyly.";
 RL Mol. Cell. Probes 8:325-327(1994).
 RN [24]
 RP VARIANTS MFS GLY-217; ASN-1023; ARG-1074; TYR-1242; ARG-1513;
 RP GLU-2127; TRP-2151; LYS-2447 AND ARG-2511.
 RP MEDLINE=94184368; PubMed=8136837;
 RA Kainulainen K., Karttunen L., Puhakka L., Sakai L., Peltonen L.;
 RT "Mutations in the fibrillin gene responsible for dominant ectopia
 RT lentis and neonatal Marfan syndrome.";
 RL Nat. Genet. 6:64-69(1994).
 RN [25]
 Query Match 12.5%; Score 383.5; DB 1; Length 2871;
 Best Local Similarity 32.6%; Pred. No. 9.5e-19;
 Matches 87; Conservative 33; Mismatches 82; Indels 65; Gaps 11;
 Qy 39 GVCHYGTKLACCYGWRNRNSKVCCEATCEP-----CKFGRCV- 75
 Db 1127 GVCH-----NTEGSRCECPFHQSPNISCIDINECELSAHLCPNRCVN 1173
 Qy 76 --GENKRCPPGY---TGKTCSDQVNECGMKPRCQHRVCNTRGSKYKFCLSGHLMPDA 130
 Db 1174 LGKYCACNPGYHSTPDLFCVDECSIMNGCCTFCTNSGSEYSCSQCPGFALMPD- 1232
 Qy 131 TCVNSRTCAINCOYSCDETE---EGPQ-----CLCPSSGLRLANGRCLDIDEC 178
 Db 1233 ----QRSCTDID---ECEDNFNIDCGGQCTNIPGEYRCLC-YDGFMASEDMKTCVDVNEC 1284

QY 179 ASKVILCPYRRVNTFGSYKCHTGFELQYISGRYDCIDINECTMDSHGTCSHHANCEN 238
 Db 1285 DLNPNIC-LSGTCENTKGSFICHDMGYSGK--KGKGTCTDINECBGAHNGCKHACVTN 1341
 QY 239 TGSFKKCKQGYKGNGLRCSAIPENS 265
 Db 1342 TAGSFKCSGPGWIGDKICTDLDECS 1368

RESULT 7

FBL2 MOUSE
 ID FBL2 MOUSE STANDARD; PRT; 1221 AA.
 AC P37889; O9MUI2;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Fibulin-2 precursor.
 GN FBLN2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 27-35.
 RC TISSUE=Fibroblast;
 RX MEDLINE=94064787; PubMed=8245130;
 RA Pan T.-C., Sasaki T., Zhang R.-Z., Faessler R., Timpl R., Chu M.-L.;
 RT "Structure and expression of fibulin-2, a novel extracellular matrix
 RT protein with multiple EGF-like repeats and consensus motifs for
 RT calcium binding.";
 RL J. Cell Biol. 123:1269-1277 (1993).
 RN [2]
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
 RX MEDLINE=99337686; PubMed=10406956;
 RA Graessel S., Sicot F.-X., Gotta S., Chu M.-L.;
 RT "Mouse fibulin-2 gene. Complete exon-intron organization and promoter
 RT characterization.";
 RL Eur. J. Biochem. 263:471-477 (1999).
 RN [3]
 RP DEVELOPMENTAL STAGE.
 RX MEDLINE=97003230; PubMed=8850569;
 RA Zhang H.-Y., Timpl R., Sasaki T., Chu M.-L., Ekblom P.;
 RT "Fibulin-1 and fibulin-2 expression during organogenesis in the
 RT developing mouse embryo.";
 RL Dev. Dyn. 205:348-364 (1996).
 RN [4]
 RP BINDING TO LAMA2.
 RX MEDLINE=99146904; PubMed=10022829;
 RA Taltis J.F., Andac Z., Goehring W., Brancaccio A., Timpl R.;
 RT "Binding of the G domains of laminin alpha1 and alpha2 chains and
 RT perlecan to heparin, sulfatides, alpha-dystroglycan and several
 RT extracellular matrix proteins.";
 RL EMBO J. 18:863-870 (1999).
 RN [5]
 RP DOWN-REGULATION BY GLUCOCORTICOID.
 RX MEDLINE=21600963; PubMed=11737251;
 RA Gu X.-C., Taltis J.F., Gullberg D., Timpl R., Ekblom M.;
 RT "Glucocorticoids down-regulate the extracellular matrix proteins
 RT fibronectin, fibulin-1 and fibulin-2 in bone marrow stroma.";
 RL Eur. J. Haematol. 67:176-184 (2001).
 CC -!- FUNCTION: Its binding to fibronectin and some other ligands is
 CC calcium dependent.
 CC -!- SUBUNIT: Homotrimer; disulfide-linked. Interacts with LAMA2.
 CC -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing; Named isoforms=2;
 CC Comment-Additional isoforms seem to exist;
 CC Name=1;
 CC IsoId=P37889-1; Sequence=Displayed;
 CC Name=2; Synonyms=EGF3-less;
 CC IsoId=P37889-2; Sequence=VSP_001391;
 CC -!- TISSUE SPECIFICITY: Component of both basement membranes and other

CC connective tissues.
 CC -!- DEVELOPMENTAL STAGE: The differential expression of the fibulin
 CC family contributes to the formation of molecularly distinct
 CC extracellular matrices already during early developmental stages
 CC of a large number of tissues.
 CC -!- INDUCTION: Glucocorticoids suppressed mRNA expression and protein
 CC synthesis.
 CC -!- SIMILARITY: Belongs to the fibulin family.
 CC -!- SIMILARITY: Contains 3 anaphylatoxin-like domains.
 CC -!- SIMILARITY: Contains 11 EGF-like domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/ebis-sib.ch/announce/>
 CC or send an email to license@ebi-sib.ch).
 CC -----
 CC EMBL; X75285; CAA53040.1; ..
 CC EMBL; AF135253; AAD34456.1; ..
 CC EMBL; AF135239; AAD34456.1; JOINED.
 CC EMBL; AF135240; AAD34456.1; JOINED.
 CC EMBL; AF135241; AAD34456.1; JOINED.
 CC EMBL; AF135242; AAD34456.1; JOINED.
 CC EMBL; AF135243; AAD34456.1; JOINED.
 CC EMBL; AF135244; AAD34456.1; JOINED.
 CC EMBL; AF135245; AAD34456.1; JOINED.
 CC EMBL; AF135246; AAD34456.1; JOINED.
 CC EMBL; AF135247; AAD34456.1; JOINED.
 CC EMBL; AF135248; AAD34456.1; JOINED.
 CC EMBL; AF135249; AAD34456.1; JOINED.
 CC EMBL; AF135250; AAD34456.1; JOINED.
 CC EMBL; AF135251; AAD34456.1; JOINED.
 CC EMBL; AF135252; AAD34456.1; JOINED.
 CC PIR; A49457; A49457.
 CC HSP; P00736; IAP0.
 CC MGD; MGI:95488; Fbln2
 CC InterPro: IPR000020; Anaphylatoxin.
 CC InterPro: IPR000152; Asx_hydroxyl_S.
 CC InterPro: IPR001881; EGF_Ca.
 CC InterPro: IPR006209; EGF_Like.
 CC Pfam; PF01821; ANATO; 2.
 CC Pfam; PF00008; EGF; 5.
 CC SMART; SM00104; ANATO; 3.
 CC SMART; SM00179; EGF_CA; 9.
 CC PROSITE; PS01177; ANAPHYLATOXIN 1; 3.
 CC PROSITE; PS01178; ANAPHYLATOXIN 2; 3.
 CC PROSITE; PS00010; ASX_HYDROXYL; 5.
 CC PROSITE; PS00022; EGF_1; FALSE_NEG.
 CC PROSITE; PS01186; EGF_2; 5.
 CC PROSITE; PS00026; EGF_3; 5.
 CC PROSITE; PS01187; EGF_CA; 10.
 KW Signal; Glycoprotein; Extracellular matrix; Plasma; EGF-like domain;
 KW Calcium-binding; Alternative splicing; Repeat.
 FT SIGNAL 1 26
 FT CHAIN 27 1221 FIBULIN-2.
 FT DOMAIN 27 434 N.
 FT DOMAIN 27 176 SUBDOMAIN NA (CYS-RICH).
 FT DOMAIN 177 434 SUBDOMAIN NB (CYS-FREE).
 FT DOMAIN 435 477 ANAPHYLATOXIN-LIKE 1.
 FT DOMAIN 478 510 ANAPHYLATOXIN-LIKE 2.
 FT DOMAIN 511 543 ANAPHYLATOXIN-LIKE 3.
 FT DOMAIN 594 635 EGF-LIKE 1, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 669 708 EGF-LIKE 2.
 FT DOMAIN 709 755 EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 756 800 EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 801 846 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 847 894 EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 895 937 EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 938 979 EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 980 1018 EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 1019 1061 EGF-LIKE 10, CALCIUM-BINDING (POTENTIAL).

between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; AF395659; AAM90567.1; --
DR InterPro; IPR000020; Anaphylatoxin.
DR InterPro; IPR000152; Asx hydrolase 1.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR006209; EGF_Like.
DR InterPro; IPR006210; IEGF.
DR Pfam; PF01821; ANATO; 1.
DR Pfam; PF00008; EGF; 4.
DR SMART; SM00104; ANATO; 1.
DR SMART; SM00181; EGF; 9.
DR SMART; SM00179; EGF_Ca; 9.
DR PROSITE; PS01177; ANAPHYLATOXIN 1; 1.
DR PROSITE; PS01178; ANAPHYLATOXIN 2; 1.
DR PROSITE; PS00010; ASX HYDROXYL; 4.
DR PROSITE; PS01186; EGF_2; 3.
DR PROSITE; PS0026; EGF_3; 4.
DR PROSITE; PS01187; EGF_Ca; 7.
KW Glycoprotein; Extracellular matrix; Repeat; EGF-like domain;
KW Calcium-binding.
FT CHAIN 1 1
FT NON_TER 1 1
FT CHAIN <1 598 FIBULIN-1C.
FT DOMAIN <1 27 ANAPHYLATOXIN-LIKE 2.
FT DOMAIN 28 60 ANAPHYLATOXIN-LIKE 3.
FT DOMAIN 92 131 EGF-LIKE 1.
FT DOMAIN 132 177 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 178 223 EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 224 270 EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 271 313 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 314 355 EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 356 395 EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 396 439 EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 440 484 EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 271 355 SELF-ASSOCIATION AND FN1-BINDING (BY
SIMILARITY).
FT DISULFID <1 25 BY SIMILARITY.
FT DISULFID 7 26 BY SIMILARITY.
FT DISULFID 28 52 BY SIMILARITY.
FT DISULFID 29 59 BY SIMILARITY.
FT DISULFID 42 60 BY SIMILARITY.
FT DISULFID 96 106 BY SIMILARITY.
FT DISULFID 102 115 BY SIMILARITY.
FT DISULFID 117 130 BY SIMILARITY.
FT DISULFID 136 149 BY SIMILARITY.
FT DISULFID 143 158 BY SIMILARITY.
FT DISULFID 164 176 BY SIMILARITY.
FT DISULFID 182 195 BY SIMILARITY.
FT DISULFID 189 204 BY SIMILARITY.
FT DISULFID 210 222 BY SIMILARITY.
FT DISULFID 228 242 BY SIMILARITY.
FT DISULFID 257 270 BY SIMILARITY.
FT DISULFID 275 288 BY SIMILARITY.
FT DISULFID 282 297 BY SIMILARITY.
FT DISULFID 299 312 BY SIMILARITY.
FT DISULFID 318 330 BY SIMILARITY.
FT DISULFID 326 339 BY SIMILARITY.
FT DISULFID 341 354 BY SIMILARITY.
FT DISULFID 360 369 BY SIMILARITY.
FT DISULFID 365 378 BY SIMILARITY.
FT DISULFID 380 394 BY SIMILARITY.
FT DISULFID 400 413 BY SIMILARITY.
FT DISULFID 409 422 BY SIMILARITY.
FT DISULFID 424 438 BY SIMILARITY.
FT DISULFID 444 457 BY SIMILARITY.
FT DISULFID 451 465 BY SIMILARITY.
FT DISULFID 471 483 BY SIMILARITY.
FT CARBOHYD 14 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 450 450 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 454 454 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 598 AA; 65516 MW; 849BF018DF452B02 CRC64;
Query Match 11.5%; Score 352.5; DB 1; Length 598;
Best Local Similarity 28.8%; Pred. No. 2.3e-17;
Matches 107; Conservative 39; Mismatches 104; Indels 121; Gaps 22;
QY 60 VCEAT-----CEP--GCKFG-----ECVGNPKRCF---PGYTGKTCQ----- 93
DB 194 ICQNTLGSFRCPKQCKNGFIQDALANCIINECLISVSAPCFIHTCINTEGSYTQKN 253
QY 94 -----DYNCGMKPRPC--QHRCVNTHSGYKCFCLSGHMLMPDATCVN 134
DB 254 VPCNGRGVHLEBGRTRCDVNECAPPECGKGRHRCVNSPGSFRCECKTYF--DGI--- 308
QY 135 SRTCAWIN-----COYSCDETEEGPQCLPSSGLRLAPNGRDCIDIDECASGKVIC 185
DB 309 SRMCDVNECORYPQRLCGHKCENTLGSYVCS--SVGRSLSDVGRSCDINECSSS---- 363
QY 186 PYNRRCVNTFGSYCKCHIGFELQVIGRYCIDINECTMDS--HTCSHHANCFNTQGSF 243
DB 364 PCSQECANVGSYQYCRGYQLSDVDG-VTCEDIDECALPTGGHICSY--RCINIPGSF 420
QY 244 KQKC-KQYK-----NGLRCSAI--PENSKEVL 270
DB 421 CCSCPASGRLAPNGRNCQDIDECVTGIHNGSINETCFNIOGGRCLAFCEPEN-----YR 476
QY 271 RAPGTIKDIRIKLLAHKNSMKKAKIKNTV-----PEPTRTPKV-----NLQPNYE 319
DB 477 RSAAT---RCERLPCHENRECKLPLR-ITYHLSFPTNIQAPVVRFMGSSAVPGDSM 532
QY 320 EIVSRGNSHG 330
DB 533 QLAIITGNEEG 543
RESULT 10
FBL1 CHICK
ID FBL1 CHICK STANDARD; PRT; 704 AA.
AC O73775; O73774;
DT 16-OCT-2001 (Rel. 40, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
GN FBLIN-1 precursor.
GN FBLIN1.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS C AND D).
RC TISSUE=Embryo;
RX MEDLINE=99120531; PubMed=9923656;
RA Barth J.L., Argaves K.M., Roark E.F., Little C.D., Argaves W.S.;
RT "Identification of chicken and C. elegans fibulin-1 homologs and
characterization of the C. elegans fibulin-1 gene";
RL Matrix Biol. 17:635-646(1998).
CC -!- FUNCTION: Incorporated into fibronectin-containing matrix fibers.
CC May play a role in cell adhesion and migration along protein
CC fibers within the extracellular matrix (ECM). Could be important
CC for certain developmental processes and contribute to the
CC supramolecular organization of ECM architecture, in particular to
CC those of basement membranes.
CC -!- SUBUNIT: Interacts with itself and with various extracellular
CC matrix components (By similarity).
CC -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=D;
CC IsoId=O73775-2; Sequence=Displayed;
CC Name=C;

SEQUENCE FROM N.A.
 STRAIN-BALB/C; TISSUE=Limb;
 MEDLINE=97238863; PubMed=9083061;
 Deak F., Piecha D., Bachrati C., Paulsson M., Kiss I.;
 "Primary structure and expression of matrilin-2, the closest relative
 of cartilage matrix protein within the von Willebrand factor type A-
 like module superfamily";
 J. Biol. Chem. 272:9268-9274(1997).
 CC - FUNCTION: Involved in matrix assembly (By similarity).
 CC - SUBCELLULAR LOCATION: Secreted
 CC - TISSUE SPECIFICITY: Detected in a variety of organs, including
 CC calvaria, uterus, heart and brain, as well as fibroblast and
 CC osteoblast cell lines.
 CC - SIMILARITY: Contains 10 EGF-like domains.
 CC - SIMILARITY: Contains 2 WFA domains.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U69262; AAC53163.1; -.
 CC HSSP; P05099; 1A05.
 CC MG; MG1:109613; Matn2.
 CC GO: GO:005578; C:extracellular matrix; IDA.
 CC InterPro; IPR000152; Asx hydroxyl_S.
 CC InterPro; IPR001881; EGF_Ca.
 CC InterPro; IPR006209; EGF_Like.
 CC InterPro; IPR006210; IEGF.
 CC InterPro; IPR002035; WFA_A.
 CC Pfam; PF00008; EGF; 10.
 CC Pfam; PF00092; WFA; 2.
 CC PRINTS; PR00453; WFADOMAIN.
 CC SMART; SM00181; EGF; 10.
 CC SMART; SM00327; WFA; 2.
 CC PROSITE; PS00010; ASX HYDROXYL; 7.
 CC PROSITE; PS01186; EGF_2; 9.
 CC PROSITE; PS0234; WFA; 2.
 CC EGF-like domain; Signal; Glycoprotein; Repeat; Coiled coil.
 CC SIGNAL
 CC CHAIN 24 956
 CC DOMAIN 57 232
 CC DOMAIN 238 278
 CC DOMAIN 279 319
 CC DOMAIN 320 360
 CC DOMAIN 361 401
 CC DOMAIN 402 442
 CC DOMAIN 443 483
 CC DOMAIN 484 524
 CC DOMAIN 525 565
 CC DOMAIN 566 606
 CC DOMAIN 607 647
 CC DOMAIN 655 830
 CC DOMAIN 917 955
 CC DISULFID 242 253
 CC DISULFID 249 262
 CC DISULFID 264 277
 CC DISULFID 283 294
 CC DISULFID 290 303
 CC DISULFID 305 318
 CC DISULFID 324 335
 CC DISULFID 331 344
 CC DISULFID 346 359
 CC DISULFID 365 376
 CC DISULFID 372 385
 CC DISULFID 387 400
 CC DISULFID 406 417
 CC DISULFID 413 426
 CC DISULFID 428 441
 CC DISULFID 447 458

FT DISULFID 454 467 BY SIMILARITY.
 FT DISULFID 469 482 BY SIMILARITY.
 FT DISULFID 488 499 BY SIMILARITY.
 FT DISULFID 495 508 BY SIMILARITY.
 FT DISULFID 510 523 BY SIMILARITY.
 FT DISULFID 529 540 BY SIMILARITY.
 FT DISULFID 536 549 BY SIMILARITY.
 FT DISULFID 551 564 BY SIMILARITY.
 FT DISULFID 570 581 BY SIMILARITY.
 FT DISULFID 577 590 BY SIMILARITY.
 FT DISULFID 592 605 BY SIMILARITY.
 FT DISULFID 611 622 BY SIMILARITY.
 FT DISULFID 618 631 BY SIMILARITY.
 FT DISULFID 633 646 BY SIMILARITY.
 FT CARBOHYD 221 221 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 890 890 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 956 AA; 106779 MW; 3E4C22770B656EEE CRC64;
 Query Match 11.1%; Score 339; DB 1; Length 956;
 Best Local Similarity 23.2%; Pred. No. 3.5e-16;
 Matches 140; Conservative 77; Mismatches 225; Indels 162; Gaps 28;
 QY 25 SARHGL---LASARQGVGHYGTKLACYWRN-SKGVCE-----ATCBGCKFGCV 75
 DB 366 ASSNHGQCHECVNAQTSALCR-----CLKGFMNPDRTKRRINYCALNKGCCSH-ECV 418
 QY 76 GPNK---CRCPGY---TGKTCODVNECGMKPRCQHRVNTGSHGKCFCLSGHMLMP 128
 DB 419 NTEGHYCRQGYNLDNGKTCR-VHCAQDGHGCLQCLNTEESFVQCQSEGLIND 477
 QY 129 D-ATCVNSRTAMIN--CQYSCDETEGPQLCPSSGLRLAPNGRDCIDIDECAGKVIC 185
 DB 478 DLKTCRADYCLLSNHGCEYSCVNTDKSPACQP-EGHVLRSDDGTCKALDSCALGDHGC 536
 QY 186 PYN-----RR-----CVNTEGSHYCKCHTGF 206
 DB 537 EHSVCSSSEDSFVCQCFEGYILRDDGKTCRRKDVQODVNHGCHLVCNSESITVCKLEGF 596
 QY 207 ELQYISGRYDCIDINECTMDSHTCSHANCFTQGSFKCKCKQGY--KNGLRCSAIPEN 264
 DB 597 RLAEQDKR--CRKNVCKSTQHCCH--MCVNGNSYLRCSEGGFVLAEDGKHCKCTEG 652
 QY 265 SVKEVLRAPTGIKRLAHKNNKKAKIKNTPEPTRP-TPKNLOPFNVE---E 320
 DB 653 PIDLVFVVDGS-----KSLGEENFETVKHFTGIIDSLAVSPKAARVGLLOYSTQVRTE 706
 QY 321 IVSRGNSHGGKKEEKKKEGLEDEKKEEK-----ALKNDIEERSLRGDFVFPKVNKA 374
 DB 707 FTLRGFSS-----AKEMKAVTHMYKMGKSGMTGLAKHMFERSFTQVEGARPPSTQV 759
 QY 375 GEGGLILVQ-----RKALTSKLEH-----KDLNISVD 401
 DB 760 PRVAIVFTDGRAQDDVSEWASAKANGITMAYGVGKALEEELQEIASEPIDKHLFYAED 819
 QY 402 CS-----FNHGICDWKQD---REDDFDNPAORDNAGFYMAVPALAGHKDGLRLK 450
 DB 820 FSTWGEISEKLEGEICEALEDSGGQDSAAWD-----LPQQAHOFTPEPT 866
 QY 451 LLLPDLPQOSNCL-----LFDYRLAGDKVGLRVFVKNNSNNALAWKEKTTSEDEKWKTKI 506
 DB 867 IKIKDLLSCSNFAVQHRFLFEEEDNLSRSTQKLFHSTKSSGNPL-----ESQDQCKCENL 921
 QY 507 QLYQ 510
 DB 922 ILFQ 925

RESULT 12
 PBL1_HUMAN STANDARD; PRT: 703 AA.
 ID PBL1_HUMAN AC P23142; P23144; P37888; Q8TBH8; Q8HBQ5; Q9UCR4; Q9U41;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)
GN Fibulin-1 precursor.
DE FBLN1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
ON NCBI_TaxID=9606;
RX [1]
RP SEQUENCE FROM N.A. (ISOFORMS A; B AND C).
RX MEDLINE=91100426; PubMed=2269669;
RA Argaves W.S., Tran H., Burgess W.H., Dickerson K.;
RT "Fibulin is an extracellular matrix and plasma glycoprotein with
RT repeated domain structure";
RL J. Cell Biol. 111:3155-3164 (1990).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM D), TISSUE SPECIFICITY, AND INTERACTION
RP WITH FN1 AND FGB.
RX MEDLINE=97260060; PubMed=9106159;
RA Tran H., Mattei M., Godyna S., Argaves W.S.;
RT "Human fibulin-1D: molecular cloning, expression and similarity with
RT SI-5 protein, a new member of the fibulin gene family.";
RL Matrix Biol. 15:479-493 (1997).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM D).
RX MEDLINE=99253993; PubMed=10318851;
RA Krachevsky A.M., Metzger E., Rosen H.;
RT "Translational control of specific genes during differentiation of
RT HL-60 cells";
RL J. Biol. Chem. 274:14295-14305 (1999).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM C).
RA Gu J.R., Wan D.F., Zhao X.T., Zhou X.M., Jiang H.O., Zhang P.P.,
RA Qin W.X., Huang Y., Qiu X.K., Qian L.F., He L.P., Li H.N., Yu Y.,
RA Yu J., Han L.H.;
RT "Novel Human cDNA clones with function of inhibiting cancer cell
RT growth";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE=20057165; PubMed=10591208;
RA Dunham I., Hunt A.R., Collins J.E., Bruskewich R., Beare D.M.,
RA Clamp M., Smink L.J., Ainscough R., Almeida J.P., Babbage A.K.,
RA Bagguley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,
RA Bird C.P., Blakey S.E., Bridgeman A.M., Buck D., Burgess J.,
RA Burrill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,
RA Clegg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R.,
RA Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,
RA Dhani P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,
RA Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,
RA Gilbert J.G.R., Goward M.E., Grahame D.V., Griffiths M.N.D., Hall C.,
RA Hall R.E., Hall-Tamlyn G., Heathcote R.W., Ho S., Holmes S.,
RA Hunt S.E., Jones M.C., Kershaw J., Kimberley A.M., King A.,
RA Laird G.K., Langford C.F., Leverisha M.A., Lloyd C., Lloyd D.M.,
RA Martyn I.D., Maffreghi-Mohammadi M., Matthews L.H., McCann O.T.,
RA McClay J., McLaren S., McMurray A.A., Milne S.A., Mortimore B.J.,
RA Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,
RA Phillips S.H., Plumb R.W., Ramsay H., Ramsey Y., Rogers L., Ross M.T.,
RA Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,
RA Soderlund C., Spragon L., Steward C.A., Sulston J.E., Swann R.M.,
RA Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,
RA Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,
RA Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,
RA Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,
RA Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyama S.,
RA Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,
RA Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.I.,
RA Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E., Nguyen T., Pan H.,
RA Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S., Sloan D., Song L.,
RA Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,
RA Zhan M., Zhang G., Chisoso S., Murray J., Miller N., Minx P.,
RA Fulton R., Johnson D., Benis G., Bentley D., Bradshaw H., Bourne S.,
RA Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,
RA Hinds K., Kemp K., Latreille P., Layman D., Ozerky P., Rohlfing T.,

RA Sheet P., Walker C., Wamsley A., Wohldmann P., Pepin K., Nelson J.,
RA Korfi I., Bedell J.A., Hillier L.W., Mardis E., Waterston R.,
RA Wilson R., Emanuel B.S., Shaikh T., Kurahashi H., Saitta S.,
RA Budarf M.L., McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E.,
RA Edelman L., Kim U.J., Shizuya H., Simon M.I., Dumanski J.P.,
RA Peyrard M., Kedra D., Serousi E., Fransson I., Tapia I., Bruder C.E.,
RA O'Brien K.P., Wilkinson P., Bodenteich A., Hartman K., Hu X.,
RA Khan A.S., Lane L., Tilahun Y., Wright H.;
RT "The DNA sequence of human chromosome 22";
RL Nature 402:489-495 (1999).
RN [6]
RP SEQUENCE FROM N.A. (ISOFORM C).
RC TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Sapleton M., Soares M.B., Bonaldo M.F., Casavant T.D., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Halle S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skaleka U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [7]
RP SEQUENCE OF 1-26 FROM N.A.
RX MEDLINE=21818451; PubMed=11829738;
RA Castoldi M., Chu M.-L.;
RT "Structural and functional characterization of the human and mouse
RT fibulin-1 gene promoters: role of Sp1 and Sp3";
RL Biochem. J. 362:41-50 (2002).
RN [8]
RP SEQUENCE OF 30-44.
RX MEDLINE=89354537; PubMed=2527614;
RA Argaves W.S., Dickerson K., Burgess W.H., Ruoslahti E.;
RT "Fibulin, a novel protein that interacts with the fibronectin
RT receptor beta subunit cytoplasmic domain";
RL Cell 58:623-629 (1989).
RN [9]
RP SELF-ASSOCIATION AND INTERACTION WITH FN1.
RX MEDLINE=93015879; PubMed=1400330;
RA Balbona K., Tran H., Godyna S., Ingham K.C., Strickland D.K.,
RA Argaves W.S.;
RT "Fibulin binds to itself and to the carboxyl-terminal heparin-binding
RT region of fibronectin";
RL J. Biol. Chem. 267:20120-20125 (1992).
RN [10]
RP POSSIBLE FUNCTION.
RX MEDLINE=95204882; PubMed=7534784;
RA Roark E.F., Keene D.R., Haudenschild C.C., Godyna S., Little C.D.,
RA Argaves W.S.;
RT "The association of human fibulin-1 with elastic fibers: an
RT immunohistochemical, ultrastructural, and RNA study";
RL J. Histochem. Cytochem. 43:401-411 (1995).
RN [11]
RP INTERACTION WITH FGB.
RX MEDLINE=95370284; PubMed=7642629;
RA Tran H., Tanaka A., Litvinovich S.V., Medved L.V., Haudenschild C.C.,
RA Argaves W.S.;
RT "The interaction of fibulin-1 with fibrinogen. A potential role in
RT hemostasis and thrombosis";
RL J. Biol. Chem. 270:19458-19464 (1995).
RN [12]

RP DEVELOPMENTAL STAGE.
RX MEDLINE=96301678; PubMed=8737292;
RA Miosge N., Gotz W., Sasaki T., Chu M.-L., Timpl R., Herken R.;
RT "The extracellular matrix proteins fibulin-1 and fibulin-2 in the
RL early human embryo.";
RN Histochim. J. 28:109-116(1996).
RP INDUCTION.
RX MEDLINE=96133928; PubMed=8552629;
RA Clinton G.M., Rougeot C., Daracourt J., Roger P., Defrenne A.,
RA Godyna S., Argraves W.S., Rochefort H.;
RT Estrogens increase the expression of fibulin-1, an extracellular
RT matrix protein secreted by human ovarian cancer cells.;
RL Proc. Natl. Acad. Sci. U.S.A. 93:316-320(1996).
RN [14]
RP CALCITRIUM, SELF-ASSOCIATION, AND FN1-BINDING SITES.
RX MEDLINE=97426402; PubMed=9278415;
RA Tran H., VanDusen W.J., Argraves W.S.;
RT "The self-association and fibronectin-binding sites of fibulin-1 map
RT to calcium-binding epidermal growth factor-like domains.";
RN J. Biol. Chem. 272:22600-22606(1997).
RN [15]
RP ROLE IN TUMOR FORMATION AND INVASION.
RX MEDLINE=98054123; PubMed=9393974;
RA Qing J., Maher V.M., Tran H., Argraves W.S., Dunstan R.W.,
RA McCormick J.J.;
RT "Suppression of anchorage-independent growth and matrigel invasion and
RT delayed tumor formation by elevated expression of fibulin-1D in human
RT fibrosarcoma-derived cell lines.";
RL Oncogene 15:2159-2168(1997).
RN [16]
RP INDUCTION.
RX MEDLINE=99027489; PubMed=9811350;
RA Roger P., Pujol P., Lucas A., Balget P., Rochefort H.;
RT "Increased immunostaining of fibulin-1, an estrogen-regulated protein
RT in the stroma of human ovarian epithelial tumors.";
RL Am. J. Pathol. 133:1579-1588(1998).
RN [17]
RP ROLE IN TUMOR FORMATION AND INVASION.
RX MEDLINE=98126147; PubMed=9466671;
RA Hayashido Y., Lucas A., Rougeot C., Godyna S., Argraves W.S.,
RA Rochefort H.;
RT "Estradiol and fibulin-1 inhibit motility of human ovarian- and
RT breast-cancer cells induced by fibronectin.";
RL Int. J. Cancer 75:654-658(1998).
RN [18]
RP INTERACTION WITH NOV.
RX MEDLINE=99128329; PubMed=927660;
RA Perbal B., Martinierie C., Sainson R., Werner M., He B., Roizman B.;
RT "The C-terminal domain of the regulatory protein NOVH is sufficient to
Query Match 11.0%; Score 338; DB 1; Length 703;
Best Local Similarity 34.9%; Pred. No. 2.8e-16;
Matches 88; Conservative 30; Mismatches 86; Indels 48; Gaps 15;
QY 46 KLACCYGWRNRKSGVC-----EATCEPCKGEGCVGNKCR-----CPFGY----T 87
DB 291 KLOCKSGFIQALGNCIDINECLISAPCPICHTINTEGVSVCQKVPNCGRGVHLMEE 350
QY 88 GKTCQSDVNECGMKPRPC--OHRCVNTHGSKVCFCLSGHMLMPDATCVNSRTCAMIN--- 142
DB 351 GTRC--VDVDECAAPPAEPCGKHRCVNSFGSFCECKTGYF--DGI-----SRMVDVNECQ 404
QY 143 -----CQYSCBDETEGFCQLCPSSGLRLAPNGRDCLDIDECASGKVICPNRRNCVTFG 196
DB 405 RYPGRLCGHKCENTLGSYLCS--SVGFRLSVDGRSCDINECSSS----PCSQECANVYG 459
QY 197 SYICKCHIGFELQYISGRYDCIDINECTWDS--HTCSHANCFTNQGFKKC--KQGYK- 252
DB 460 SYQYCCRYGQLSDVDG--VTCEIDICALPTGHTICSY--RCINIPGSGFQSCPSGVR 516
QY 253 -GNGLRCSAIP 263

DB 517 APNGRNCQDIDE 528

RESULT 13
NOTC DROME
ID NOTC DROME STANDARD; PRT: 2703 AA.
AC P07207; O97458; P04154; Q9W4T8;
DT 01-NOV-1986 (Rel. 03, Created)
DT 28-PEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Neurogenic locus Notch protein precursor.
GN N OR EG:140G11.1 OR EG:163A10.2 OR CG3936.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Oregon-R; TISSUE=Embryo;
RX MEDLINE=86079539; PubMed=3935325;
RA Wharton K.A., Johansen K.M., Xu T., Artavanis-Tsakonas S.;
RT "Nucleotide sequence from the neurogenic locus notch implies a gene
RT product that shares homology with proteins containing EGF-like
RT repeats.";
RL Cell 43:567-581(1985).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Canton-S, and Oregon-R; TISSUE=Embryo;
RX MEDLINE=87064624; PubMed=3097517;
RA Kidd S., Kelley M.R., Young M.W.;
RT "Sequence of the notch locus of Drosophila melanogaster: relationship
RL of the encoded protein to mammalian clotting and growth factors.";
RN Mol. Cell. Biol. 6:3094-3108(1986).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA Sutton G., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abail J.F., Agbavani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Balles R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Foler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.I., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy E., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun B.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,


```

Db 270 CRNGSLYECKPKPGFEKGNCEQNYDDCLGHLQNGGTCTDGSIDYTCRCPPNFTGRFCQ 329
QY 174 -DIDECAS-----GKVICPYNRRCVN 193
Db 330 DDVDECAQRHPVQCGATCTNTHGYSVCIVNGWAGLDCSNNTDDCKQACFYGATCID 389
QY 194 TFGSYCKCHIG-----FELOYTSGRYDCI-----D 219
Db 390 GVGSGFYCQCTKGKTGLLCHLDDACTSNPCHADAICTDPTSPINGSYACSCATGYKGVDCSED 449
QY 220 INECTMDSHTCSHANCENTQSGFKCKCKGYKG-----NGLRCSALPENSVEKVLAPG 274
Db 450 IDECDQGS-PCHNGICVNTPGSYRCNCSQOFTGPRCETNINECESHPQNEGSLDDPG 508
QY 275 TIK 277
Db 509 TFR 511

RESULT 14
FBL1_MOUSE STANDARD; PRT; 705 AA.
AC Q08879; Q08878; Q8C3B1; Q91ZG9; Q922K8;
DT 01-OCT-1994 (Rel. 30, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Fibulin-1 precursor (Basement-membrane protein 90) (BM-90).
GN FBLN1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A. (ISOFORMS C AND D), AND LIGANDS INTERACTION.
RX MEDLINE=93358897; PubMed=8354280;
RA Pan T.-C., Kluge M., Zhang R.Z., Mayer U., Timpl R., Chu M.-L.;
RT "Sequence of extracellular mouse protein BM-90/fibulin and its
RT calcium-dependent binding to other basement-membrane ligands.";
RL Eur. J. Biochem. 215:733-740 (1993).
[2]
SEQUENCE FROM N.A. (ISOFORM C).
RX STRAIN=C57BL/6J; TISSUE=Head, and Urinary bladder;
RX MEDLINE=22354683; PubMed=12466851;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA Nikaïdo I., Osato N., Sato R., Suzuki H., Imanaka I., Kiyosawa H.,
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA Dalla E., Dragani T.A., Fletcher V., Clothia C., Corbani L.E., Cousins S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Frazer K.S.,
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA Maglott D.R., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA Ravasi T., Reed J.C., Reid J.J., Ring B.Z., Ringwald M.,
RA Sandelin A., Schneider C., Sempile C.A., Setou M., Shimada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA Wilming L.G., Wyszynski-Boris A., Yanagisawa M., Yang I., Yang L.,
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA Hara A., Hashizume W., Inotani K., Ishii Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shingawa A.,
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA Birney E., Hayashizaki Y.;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).

```

```

RN RP SEQUENCE FROM N.A. (ISOFORM C).
RC TISSUE=Breast tumor;
RX MEDLINE=22388257; PubMed=12477932;
RA Srausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Sosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RL and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
[4]
SEQUENCE OF 1-26 FROM N.A.
RX MEDLINE=21818451; PubMed=11829738;
RA Castoldi M., Chu M.-L.;
RT "Structural and functional characterization of the human and mouse
RL fibulin-1 gene promoters: role of Sp1 and Sp3.";
RL Biochem. J. 362:41-50 (2002).
[5]
CHARACTERIZATION OF NID AFFINITY.
RX MEDLINE=95147264; PubMed=7844816;
RA Sasaki T., Kostka G., Goehring W., Wiedemann H., Mann K., Chu M.-L.,
RA Timpl R.;
RT "Structural characterization of two variants of fibulin-1 that differ
RL in nitrogen affinity.";
RL J. Mol. Biol. 245:241-250 (1995).
[6]
DEVELOPMENTAL STAGE.
RX MEDLINE=97003230; PubMed=8850569;
RA Zhang H.-Y., Timpl R., Sasaki T., Chu M.-L., Ekblom P.;
RT "Fibulin-1 and fibulin-2 expression during organogenesis in the
RL developing mouse embryo.";
RL Dev. Dyn. 205:348-364 (1996).
[7]
NID-BINDING SITE.
RX STRAIN=129/Sv.
RX MEDLINE=97446166; PubMed=9299350;
RA Adam S., Goehring W., Wiedemann H., Chu M.-L., Timpl R., Kostka G.;
RT "Binding of fibulin-1 to nidogen depends on its C-terminal globular
RL domain and a specific array of calcium-binding epidermal growth
RL factor-like (EG) modules.";
RL J. Mol. Biol. 272:226-236 (1997).
[8]
BINDING TO LAMA2.
RX MEDLINE=99146904; PubMed=10022829;
RA Tait J.F., Andac Z., Goehring W., Brancaccio A., Timpl R.;
RT "Binding of the G domains of laminin alpha1 and alpha2 chains and
RL perlecan to heparin, sulfatides, alpha-dystroglycan and several
RL extracellular matrix proteins.";
RL EMBO J. 18:863-870 (1999).
[9]
INTERACTION WITH AGC1 AND CSPG2.
RX MEDLINE=99329059; PubMed=10400671;
RA Asperger A., Adam S., Kostka G., Timpl R., Heinegaard D.;
RT "Fibulin-1 is a ligand for the C-type lectin domains of aggrecan and
RL versican.";
RL J. Biol. Chem. 274:20444-20449 (1999).
[10]
INTERACTION WITH NID.
RX MEDLINE=21474010; PubMed=11589703;

```

RA Ries A., Goehring W., Fox J.W., Timpl R., Sasaki T.;
 RT "Recombinant domains of mouse nidogen-1 and their binding to basement
 RL membrane proteins and monoclonal antibodies.";
 RN Eur. J. Biochem. 268:5119-5128(2001).
 [11]
 RN DOWN-REGULATION BY GLUCOCORTICOID.
 RP MEDLINE=21600963; PubMed=11737251;
 RX MEDLINE=21600963; PubMed=11737251;
 RA Gu Y.-C., Tals J.F., Gullberg D., Timpl R., Ekblom M.;
 RT "Glucocorticoids down-regulate the extracellular matrix proteins
 RL fibronectin, fibulin-1 and fibulin-2 in bone marrow stroma.";
 RN Eur. J. Haematol. 67:176-184(2001).
 [12]
 RN TISSUE SPECIFICITY, AND DEVELOPMENTAL STAGE.
 RP MEDLINE=21136579; PubMed=11238726;
 RX MEDLINE=21136579; PubMed=11238726;
 RA Ohsawa I., Takamura C., Kohsaka S.;
 RT "Fibulin-1 binds the amino-terminal head of beta-amyloid precursor
 RL protein and modulates its physiological function.";
 RN J. Neurochem. 76:1411-1420(2001).
 [13]
 RP INTERACTION WITH E6, AND INHIBITION OF E6-MEDIATED TRANSFORMATION.
 RX MEDLINE=22188366; PubMed=12200142;
 RA Du M., Fan X., Hong E., Chen J.J.;
 RT "Interaction of oncogenic papillomavirus E6 proteins with fibulin-1.";
 RL Biochem. Biophys. Res. Commun. 296:962-969(2002).
 [14]
 RN DEVELOPMENTAL STAGE.
 RP MEDLINE=21826618; PubMed=11836357;
 RX MEDLINE=21826618; PubMed=11836357;
 RA Debeer P., Schoenmakers E.F.P.M., Twa W.O., Argraves W.S.,
 RT De Smet L., Frys J.-P., Van De Ven W.J.M.;
 RL "The fibulin-1 gene (FBLN1) is disrupted in a t(12;22) associated with
 RT a complex type of synpolydactyly.";
 RN J. Med. Genet. 39:98-104(2002).
 [15]
 CC -!- FUNCTION: Incorporated into fibronectin-containing matrix fibers.
 CC May play a role in cell adhesion and migration along protein
 CC fibers within the extracellular matrix (ECM). Could be important
 CC for certain developmental processes and contribute to the
 CC supramolecular organization of ECM architecture, in particular to
 CC those of basement membranes.
 CC -!- SUBUNIT: Interacts with itself and with various extracellular
 CC matrix components such as FN1, LAMAL, LMA2, NID, AGC1, CSPG2 and
 CC type IV collagen. Interacts also with papillomavirus E6 proteins.
 CC Binding analysis demonstrated for isoform C a 100-fold stronger
 CC binding to the basement membrane protein NID than for isoform D.
 CC -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
 CC -!- ALTERNATIVE PRODUCTS:
 CC -!- Event=Alternative splicing; Named isoforms=4;
 CC Name=D;
 CC IsoId=Q08879-1; Sequence=Displayed;
 CC Name=A;
 CC IsoId=Q08879-3; Sequence=Not described;
 CC Name=B;
 CC IsoId=Q08879-4; Sequence=Not described;
 CC Name=C;
 CC IsoId=Q08879-2; Sequence=VSP 001386;
 CC Note=Ref.1 isoform C is in conflict in position(s): 571:E-A;
 CC -!- TISSUE SPECIFICITY: Detected in most organs (brain, heart, lung,
 CC spleen, liver and kidney). Neurons are the predominant source of
 CC production in the brain. Not expressed significantly by astrocytes
 CC or microglia.
 CC -!- DEVELOPMENTAL STAGE: The differential expression of the fibulin
 CC family contributes to the formation of molecularly distinct
 CC extracellular matrices already during early developmental stages
 CC of a large number of tissues. Increase expression at neonate stage
 CC in the brain. Expressed in interdigital regions of the handplate
 CC of a 12 dpc embryo and in the lateral perichondrial region.
 CC Similar expression persists in the 13 dpc handplate particularly
 CC in the perichondrial regions and apical aspects of the developing
 CC digits.
 CC -!- INDUCTION: Glucocorticoids suppressed mRNA expression and protein
 CC synthesis.
 CC -!- SIMILARITY: Belongs to the fibulin family.
 CC -!- SIMILARITY: Contains 3 anaphylatoxin-like domains.
 CC -!- SIMILARITY: Contains 9 EGF-like domains.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC -----
 CC Query Match 10.9%; Score 335; DB 1; Length 705;
 CC Best Local Similarity 34.1%; Pred. No. 4.6e-16;
 CC Matches 86; Conservative 31; Mismatches 87; Indels 48; Gaps 15;
 CC -----
 CC QY 46 KLACCYGWRNRSGVC-----EATCEGCKFGECVGNKCR-----CPGY-----T 87
 CC Db 293 KLQKSGFTQDALGNICDINECLISAPCPVGTCTINTEGTYCQKVPNCGRGYHLNEE 352
 CC QY 88 GKTCSQDVNECKMPKPC--QHRVNTVHSGYKCFCLSGHMLMPDATCVNSRTCAIN--- 142
 CC Db 353 GTRC-VDVDECSPPAEPCKGKHCLNSPSPCRCKAGYF--DGI---SRTVDINPCQ 406
 CC QY 143 -----COYSCEDTEBGPQCLPSSGLRLAPNGRDCLDIDECASGKVCYPNRRCVNFTG 196
 CC Db 407 RYFGRLCGHKCENTPGSFHCSC-SAGFRLSVDRGSCEDVNECLNS-----PCSQECANVYG 461
 CC QY 197 SYICKCHIGFELQVIGRYDCIDINECTMDS--HTCSHHANCFNTQSPKCK-KQGYK- 252
 CC Db 462 SYQCYCRGVQLSDVDG-VTCEDIDECALPTGGHICSY--RCINIPGSGFCQSPSSGYRL 518
 CC QY 253 -GNGLRCSAIP 263
 CC Db 519 APNGNQCQDIDE 530
 CC -----
 CC RESULT 15
 CC LTBS MOUSE
 CC ID LTBS MOUSE STANDARD; PRT; 1389 AA.
 CC AC OBOG18; O8BNW7; O8C7F5; O8C1R0;
 CC DT 10-OCT-2003 (Rel. 42, Created)
 CC DT 10-OCT-2003 (Rel. 42, Last sequence update)
 CC DT 10-OCT-2003 (Rel. 42, Last annotation update)
 CC DE Latent transforming growth factor beta binding protein, isoform 15
 CC DE precursor (LTBP-1) (Transforming growth factor beta-1 binding protein
 CC DE 1) (TGF-beta1-BP-1).
 CC GN LTBP1.
 CC OS Mus musculus (Mouse).
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CC OC NCBI_TaxID=10090;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RX MEDLINE=22597538; PubMed=12711388;
 CC RA Noguera I., Obata H., Gualandris A., Cowin P., Rifkin D.B.;
 CC RT "Molecular cloning of the mouse Ltbp-1 gene reveals tissue specific
 CC RT expression of alternatively spliced forms.";
 CC RL Gene 308:31-41(2003).
 CC RN [2]
 CC RP SEQUENCE FROM N.A.
 CC RX STRAIN=129/SVJ;
 CC RC MEDLINE=22597539; PubMed=12711389;
 CC RA Weiskirchen R., Moser M., Guenther K., Weiskirchen S., Gressner A.M.;
 CC RT "The murine latent transforming growth factor-beta binding protein
 CC RT (Ltbp-1) is alternatively spliced, and maps to a region syntenic to
 CC RT human chromosome 2p21-22.";
 CC RL Gene 308:43-52(2003).
 CC RN [3]
 CC RP SEQUENCE OF 788-1389 FROM N.A.
 CC RC STRAIN=C57BL/6J; TISSUE=Aorta, Liver, and Vein;
 CC RX MEDLINE=22354683; PubMed=12466851;
 CC RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
 CC RA Nikaide I., Otsu N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
 CC RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojibori T.,
 CC RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
 CC RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
 CC RA Blake J.A., Bradt D., Brusica V., Chothia C., Corbani L.E., Cousins S.,
 CC RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
 CC RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
 CC RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,

RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA Nagashima T., Nunata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA Vezardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA Birney E., Hayashizaki Y.,
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs";
RL Nature 420:563-573(2002).
CC -!- SUBUNIT: The large latent complex of TGF-beta1 from platelets is
CC composed of the TGF-beta1 molecule noncovalently associated with a
CC disulfide-bonded complex of a dimer of the N-terminal propeptide
CC of the TGF-beta1 precursor and a third component denoted TGF-
CC beta1-BP. TGF-beta1-BP does not bind directly to active TGF-BETA1.
CC Binds to fibrillin (By similarity).
CC -!- SUBCELLULAR LOCATION: Secreted (By similarity).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=IS; IsoId=Q8CG18-1; Sequence=Displayed;
CC Name=IL;
CC IsoId=Q8CG19-1; Sequence=External;
CC -!- PTM: Contains hydroxylated asparagine residues (By similarity).
CC -!- PTM: The N-terminus is blocked (By similarity).
CC -!- SIMILARITY: Contains 16 EGF-like domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AF346465; AAN77251.1; JOINED.
DR EMBL; AF346438; AAN77251.1; JOINED.
DR EMBL; AF346439; AAN77251.1; JOINED.
DR EMBL; AF346440; AAN77251.1; JOINED.
DR EMBL; AF346441; AAN77251.1; JOINED.
DR EMBL; AF346442; AAN77251.1; JOINED.
DR EMBL; AF346443; AAN77251.1; JOINED.
DR EMBL; AF346444; AAN77251.1; JOINED.
DR EMBL; AF346445; AAN77251.1; JOINED.
DR EMBL; AF346446; AAN77251.1; JOINED.
DR EMBL; AF346447; AAN77251.1; JOINED.
DR EMBL; AF346448; AAN77251.1; JOINED.
DR EMBL; AF346449; AAN77251.1; JOINED.
DR EMBL; AF346450; AAN77251.1; JOINED.
DR EMBL; AF346451; AAN77251.1; JOINED.
DR EMBL; AF346452; AAN77251.1; JOINED.
DR EMBL; AF346453; AAN77251.1; JOINED.
DR EMBL; AF346454; AAN77251.1; JOINED.
DR EMBL; AF346455; AAN77251.1; JOINED.
DR EMBL; AF346456; AAN77251.1; JOINED.
DR EMBL; AF346457; AAN77251.1; JOINED.
DR EMBL; AF346458; AAN77251.1; JOINED.
DR EMBL; AF346459; AAN77251.1; JOINED.
DR EMBL; AF346460; AAN77251.1; JOINED.
DR EMBL; AF346461; AAN77251.1; JOINED.
DR EMBL; AF346462; AAN77251.1; JOINED.
DR EMBL; AF346463; AAN77251.1; JOINED.
DR EMBL; AF346464; AAN77251.1; JOINED.

DR EMBL; AY143161; AAN38831.1; ALT_SEQ.
DR EMBL; AK050380; BAC34222.1; -.
DR EMBL; AK080024; BAC37808.1; -.
DR MGD; MGI:109151; Ldbpl.
DR GO; GO:0005578; C:extracellular matrix; IDA.
DR InterPro; IPR000152; Asx_Hydroxyl_S.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR006209; EGF_Like.
DR InterPro; IPR002212; Fibril-assoc.
DR InterPro; IPR006210; IEGF.
DR Pfam; PF00008; EGF; 15.
DR Pfam; PF00683; TB; 4.
DR SMART; SM00181; EGF; 17.
DR SMART; SM00179; EGF_Ca; 16.
DR PROSITE; PS00010; ASX_HYDROXYL; 13.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 11.
DR PROSITE; PS00026; EGF_3; 13.
DR PROSITE; PS01187; EGF_Ca; 15.
DR Growth factor binding; Repeat; EGF-like domain;
KW Hydroxylation; Signal; Glycoprotein;
KW Alternative splicing.
FT SIGNAL 1 19
FT CHAIN 20 1389
FT DOMAIN 295 335
FT REPEAT 343 408
FT DOMAIN 542 583
FT DOMAIN 584 625
FT DOMAIN 626 666
FT DOMAIN 667 706
FT DOMAIN 707 747
FT DOMAIN 748 788
FT DOMAIN 789 829
FT DOMAIN 830 870
FT DOMAIN 871 912
FT DOMAIN 913 954
FT DOMAIN 955 997
FT REPEAT 1012 1079
FT DOMAIN 1092 1134
FT DOMAIN 1135 1175
FT REPEAT 1185 1257
FT DOMAIN 1289 1330
FT DOMAIN 1330 1374
FT DISULFID 299 310
FT DISULFID 305 319
FT DISULFID 321 334
FT DISULFID 546 558
FT DISULFID 553 567
FT DISULFID 569 582
FT DISULFID 588 600
FT DISULFID 595 609
FT DISULFID 611 624
FT DISULFID 630 641
FT DISULFID 636 650
FT DISULFID 653 665
FT DISULFID 671 682
FT DISULFID 677 691
POTENTIAL
LATENT TRANSFORMING GROWTH FACTOR BETA
BINDING PROTEIN, ISOFORM 1S.
EGF-LIKE 1, CALCIUM-BINDING
(POTENTIAL).
REPEAT A.
EGF-LIKE 2, CALCIUM-BINDING
(POTENTIAL).
EGF-LIKE 3, CALCIUM-BINDING
(POTENTIAL).
EGF-LIKE 4, CALCIUM-BINDING
(POTENTIAL).
EGF-LIKE 5, CALCIUM-BINDING
(POTENTIAL).
EGF-LIKE 6, CALCIUM-BINDING
(POTENTIAL).
EGF-LIKE 7, CALCIUM-BINDING
(POTENTIAL).
EGF-LIKE 8, CALCIUM-BINDING
(POTENTIAL).
EGF-LIKE 9, CALCIUM-BINDING
(POTENTIAL).
EGF-LIKE 10, CALCIUM-BINDING
(POTENTIAL).
EGF-LIKE 11, CALCIUM-BINDING
(POTENTIAL).
EGF-LIKE 12, CALCIUM-BINDING
(POTENTIAL).
REPEAT B.
EGF-LIKE 13, CALCIUM-BINDING
(POTENTIAL).
EGF-LIKE 14, CALCIUM-BINDING
(POTENTIAL).
REPEAT C.
EGF-LIKE 15, CALCIUM-BINDING
(POTENTIAL).
POTENTIAL.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 21, 2004, 12:27:57 ; Search time 70.0986 Seconds
(without alignments)
2489.088 Million cell updates/sec

Title: US-09-981-649A-24

Perfect score: 3060

Sequence: 1 MFLPWSALPLLSWVAGF.....VDGVLLVSLGCPDLSLVDD 553

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp_arChaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3060	100.0	553	Q81UX8	Q81ux8 homo sapien
2	3055	99.8	553	Q8NBV0	Q8nbv0 homo sapien
3	3055	99.8	553	Q9NZL7	Q9nzl7 homo sapien
4	3049.5	99.7	554	Q9NY67	Q9ny67 homo sapien
5	3016	98.6	558	Q9UFK6	Q9ufk6 homo sapien
6	2541	83.0	474	Q8WYG3	Q8wys3 homo sapien
7	2401.5	78.5	550	Q9JUZ5	Q9jjz5 mus musculus
8	1630.5	53.3	544	Q8AVH7	Q8avh7 xenopus lae
9	1486.5	48.6	327	Q8BPM8	Q8bpm8 mus musculus
10	1095	35.8	561	Q91V88	Q91v88 mus musculus
11	1081.5	35.3	578	Q91ZD3	Q91zd3 mus musculus
12	1069.5	35.0	592	Q91XL5	Q91xl5 mus musculus
13	1056	34.5	609	Q923T5	Q923t5 mus musculus
14	472	15.4	372	Q80VP6	Q80vp6 mus musculus
15	408.5	13.3	2906	Q9WUH9	Q9wuh9 rattus norv
16	405	13.2	2809	Q96JP8	Q96jp8 homo sapien

17	405	13.2	2809	4	Q86SU5	Q86sj5 homo sapien
18	386	12.6	708	13	P87363	P87363 gallus gall
19	385.5	12.6	2872	11	Q9WUH8	Q9wuh8 rattus norv
20	383.5	12.5	3857	11	Q8840	Q8840 mus musculus
21	382	12.5	1174	11	Q98K58	Q98k58 mus musculus
22	374.5	12.2	2189	5	Q9BI05	Q9bi05 eimeria ten
23	370	12.1	608	11	Q80V54	Q80v54 mus musculus
24	369	12.1	937	5	Q9BLJ1	Q9blj1 ciona intes
25	365.5	11.9	1574	11	Q88281	Q88281 rattus norv
26	362.5	11.8	576	4	Q9Y3V7	Q9y3v7 homo sapien
27	362.5	11.8	1184	4	Q86V58	Q86v58 homo sapien
28	362.5	11.8	1231	4	Q8IUI1	Q8iui1 homo sapien
29	362.5	11.8	1231	4	Q8IUI0	Q8iui0 homo sapien
30	362.5	11.8	1511	4	Q75412	Q75412 homo sapien
31	361.5	11.8	1587	4	Q00508	Q00508 homo sapien
32	358	11.7	2360	5	Q7YZP0	Q7yzp0 eimeria max
33	354.5	11.6	955	4	Q96DN2	Q96dn2 homo sapien
34	354.5	11.6	1833	11	Q08999	Q08999 mus musculus
35	351	11.5	1600	11	Q8K4G0	Q8k4g0 mus musculus
36	351	11.5	1666	11	Q8K4G1	Q8k4g1 mus musculus
37	349	11.4	2673	4	Q96SC3	Q96sc3 homo sapien
38	342.5	11.2	1246	4	Q75095	Q75095 homo sapien
39	340.5	11.1	1664	5	Q9TVQ2	Q9tvq2 caenorhabdi
40	339.5	11.1	741	4	Q96K89	Q96k89 homo sapien
41	339	11.1	956	11	Q99K64	Q99k64 mus musculus
42	339	11.1	956	11	Q8R542	Q8r542 mus musculus
43	339	11.1	1764	11	Q35806	Q35806 rattus norv
44	338	11.0	528	11	Q9CXD8	Q9cxd8 mus musculus
45	338	11.0	638	4	Q8NBH6	Q8nbh6 homo sapien

ALIGNMENTS

RESULT 1

Q81UX8	
ID	Q81UX8
AC	Q81UX8; PRELIMINARY; PRT; 553 AA.
DT	01-MAR-2003 (TrEMBLrel. 23, Created)
DT	01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE	Similar to EGF-like-domain, multiple 6.
OS	Homo sapiens (Human)
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX	NCBI_TaxID=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=Brain;
RA	Strausberg R.;
RL	Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
DR	ENBL; BC038587; AAH38587.1; -
DR	GO; GO:0016020; C:membrane; IEA.
DR	GO; GO:0005509; F:calcium ion binding; IEA.
DR	InterPro; IPR000152; Asx_hydroxyl_S.
DR	InterPro; IPR001881; EGF_Ca.
DR	InterPro; IPR006209; EGF_Like.
DR	InterPro; IPR006210; IEGF.
DR	InterPro; IPR000998; MAM_domain.
DR	Pfam; PF00008; EGF; 4.
DR	Pfam; PF00629; MAM; 1.
DR	SMART; SM00181; EGF; 5.
DR	SMART; SM00179; EGF_CA; 3.
DR	SMART; SM00137; MAM; 1.
DR	PROSITE; PS00010; ASX_HYDROXYL; 3.
DR	PROSITE; PS00022; EGF_1; 1.
DR	PROSITE; PS01186; EGF_2; 2.
DR	PROSITE; PS01187; EGF_CA; 3.
DR	PROSITE; PS00060; MAM_2; 1.
SQ	SEQUENCE 553 AA; 61317 MW; 3AE93A0362E861E0 CRC64;

Query Match 100.0%; Score 3060; DB 4; Length 553;
Best Local Similarity 100.0%; Pred. No. 3.7e-233;

```

Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MPLPWSLALPILLSWVAGFGNAASARHHGLLASARQPGVCHYGTKLACCYGWRRNSKGV 60
DB 1 MPLPWSLALPILLSWVAGFGNAASARHHGLLASARQPGVCHYGTKLACCYGWRRNSKGV 60
QY 61 CEATCEPGCKFGEVGNPKRCFPFGYTGKTCSDQVNECGMKPRPCQHRVCVNTHTSGYKFC 120
DB 61 CEATCEPGCKFGEVGNPKRCFPFGYTGKTCSDQVNECGMKPRPCQHRVCVNTHTSGYKFC 120
QY 121 LSGHMLMPDATCVNSRTCAMINCQYSCEDTEEGPQCLPSSGLRLAPNGRDCLDIDECAS 180
DB 121 LSGHMLMPDATCVNSRTCAMINCQYSCEDTEEGPQCLPSSGLRLAPNGRDCLDIDECAS 180
QY 181 GKVICPNRRRCVNTFGSYCKCHIGFELYISGRYDCIDINECTMDSHTCSHANCFTQ 240
DB 181 GKVICPNRRRCVNTFGSYCKCHIGFELYISGRYDCIDINECTMDSHTCSHANCFTQ 240
QY 241 GSFKCKCKQGYKGNGLRCSAIPENSVEVLRAPGTIKDRIKLLAHKNSMKKAKIKNVT 300
DB 241 GSFKCKCKQGYKGNGLRCSAIPENSVEVLRAPGTIKDRIKLLAHKNSMKKAKIKNVT 300
QY 301 PEPTTPTPKVNLQPFNYEEIVSRGNSHGKKGNEEKMEGLEDEKREKALKNDIEER 360
DB 301 PEPTTPTPKVNLQPFNYEEIVSRGNSHGKKGNEEKMEGLEDEKREKALKNDIEER 360
QY 361 SLRGDVFPPKNEAGEFGLILVQRKALTSKLEHKDLNISVDCSFNHHGICDWKQDREDDFD 420
DB 361 SLRGDVFPPKNEAGEFGLILVQRKALTSKLEHKDLNISVDCSFNHHGICDWKQDREDDFD 420
QY 421 WNPADRDNAIGFYMAVPALAGHKDIDIGRLKLLPDLQPSNFCLLFDYRLAGDKVGLRV 480
DB 421 WNPADRDNAIGFYMAVPALAGHKDIDIGRLKLLPDLQPSNFCLLFDYRLAGDKVGLRV 480
QY 481 FVKNNSNALAWKTTSEDEKWKTKGIQLYQGTDTATKSIIFEAEERGKGTGEIADVGVLLV 540
DB 481 FVKNNSNALAWKTTSEDEKWKTKGIQLYQGTDTATKSIIFEAEERGKGTGEIADVGVLLV 540
QY 541 SGLCPDLSLLSVDD 553
DB 541 SGLCPDLSLLSVDD 553

RESULT 2
Q8NEVO PRELIMINARY; PRT; 553 AA.
AC Q8NEVO;
DT 01-OCT-2002 (TremBLrel. 22, Created)
DT 01-OCT-2002 (TremBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)
DE Hypothetical protein FLJ90733.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Isogai T., Ota T., Nishikawa T., Hayashi K., Otsuki T., Sugiyama T.,
RA Suzuki Y., Nagai K., Sugano S., Ishii S., Kawai-Hio Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Kojima S., Negahari K.,
RA Masuho Y., Ono T., Okano K., Yoshikawa Y., Aotsuka S., Sasaki N.,
RA Hattori A., Okumura K., Iwayanagi T., Ninomiya K.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: CONTAINS 1 MAM DOMAIN.
DR EMBL; AK075214; BAC11477.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR InterPro; IPR000152; ASX_hydroxyl_S.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR006210; IEGF.

```

```

InterPro; IPR000998; MAM_domain.
PFam; PF00008; EGF; 4.
PFam; PF00629; MAM; 1.
SMART; SM00181; EGF; 5.
SMART; SM00179; EGF_CA; 4.
SMART; SM00137; MAM; 1.
PROSITE; PS00010; ASX_HYDROXYL; 3.
PROSITE; PS00022; EGF_1; 1.
PROSITE; PS01186; EGF_2; 2.
PROSITE; PS01187; EGF_CA; 3.
PROSITE; PS00060; MAM_2; 1.
KW Hypothetical protein; EGF-like domain; Glycoprotein.
SQ SEQUENCE 553 AA; 61318 MW; 3AE93A013CED5880 CRC64;

Query Match 99.8%; Score 3055; DB 4; Length 553;
Best Local Similarity 99.8%; Pred. No. 9.1e-233;
Matches 552; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 MPLPWSLALPILLSWVAGFGNAASARHHGLLASARQPGVCHYGTKLACCYGWRRNSKGV 60
DB 1 MPLPWSLALPILLSWVAGFGNAASARHHGLLASARQPGVCHYGTKLACCYGWRRNSKGV 60
QY 61 CEATCEPGCKFGEVGNPKRCFPFGYTGKTCSDQVNECGMKPRPCQHRVCVNTHTSGYKFC 120
DB 61 CEATCEPGCKFGEVGNPKRCFPFGYTGKTCSDQVNECGMKPRPCQHRVCVNTHTSGYKFC 120
QY 121 LSGHMLMPDATCVNSRTCAMINCQYSCEDTEEGPQCLPSSGLRLAPNGRDCLDIDECAS 180
DB 121 LSGHMLMPDATCVNSRTCAMINCQYSCEDTEEGPQCLPSSGLRLAPNGRDCLDIDECAS 180
QY 181 GKVICPNRRRCVNTFGSYCKCHIGFELYISGRYDCIDINECTMDSHTCSHANCFTQ 240
DB 181 GKVICPNRRRCVNTFGSYCKCHIGFELYISGRYDCIDINECTMDSHTCSHANCFTQ 240
QY 241 GSFKCKCKQGYKGNGLRCSAIPENSVEVLRAPGTIKDRIKLLAHKNSMKKAKIKNVT 300
DB 241 GSFKCKCKQGYKGNGLRCSAIPENSVEVLRAPGTIKDRIKLLAHKNSMKKAKIKNVT 300
QY 301 PEPTTPTPKVNLQPFNYEEIVSRGNSHGKKGNEEKMEGLEDEKREKALKNDIEER 360
DB 301 PEPTTPTPKVNLQPFNYEEIVSRGNSHGKKGNEEKMEGLEDEKREKALKNDIEER 360
QY 361 SLRGDVFPPKNEAGEFGLILVQRKALTSKLEHKDLNISVDCSFNHHGICDWKQDREDDFD 420
DB 361 SLRGDVFPPKNEAGEFGLILVQRKALTSKLEHKDLNISVDCSFNHHGICDWKQDREDDFD 420
QY 421 WNPADRDNAIGFYMAVPALAGHKDIDIGRLKLLPDLQPSNFCLLFDYRLAGDKVGLRV 480
DB 421 WNPADRDNAIGFYMAVPALAGHKDIDIGRLKLLPDLQPSNFCLLFDYRLAGDKVGLRV 480
QY 481 FVKNNSNALAWKTTSEDEKWKTKGIQLYQGTDTATKSIIFEAEERGKGTGEIADVGVLLV 540
DB 481 FVKNNSNALAWKTTSEDEKWKTKGIQLYQGTDTATKSIIFEAEERGKGTGEIADVGVLLV 540
QY 541 SGLCPDLSLLSVDD 553
DB 541 SGLCPDLSLLSVDD 553

RESULT 3
Q9NZL7 PRELIMINARY; PRT; 553 AA.
AC Q9NZL7;
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)
DE Epidermal growth factor repeat containing protein.
GN EGF6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]

```

RP SEQUENCE FROM N.A.
EX MEDLINE=20079166; PubMed=10610727;
RA Ford J.E.;
RT "Cloning of a novel epidermal growth factor repeat containing gene
RT EGF16; expressed in tumor and fetal tissues.";
RL Genomics 62:304-307(1999).
CC -!- SIMILARITY: CONTAINS 1 MAM DOMAIN.
DR EMBL; AF186084; AAF27812.1; -.
DR HSP; P00736; IAPQ.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR InterPro; IPR000152; Asx_hydroxyl_S.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR000998; MAM_domain.
DR Pfam; PF00008; EGF; 4.
DR Pfam; PF00629; MAM; 1.
DR SMART; SM00179; EGF_CA; 3.
DR PROSITE; PS00010; ASX_HYDROXYL; 3.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 2.
DR PROSITE; PS01187; EGF_CA; 3.
DR PROSITE; PS00060; MAM_2; 1.
KW EGF-like domain; Glycoprotein.
SQ SEQUENCE 553 AA; 61314 MW; 2FF55F167857DE50 CRC64;

Query Match 99.8%; Score 3055; DB 4; Length 553;
Best Local Similarity 99.8%; Pred. No. 9.1e-233;
Matches 552; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MFLPWSLALPLLSSWAGGFGNNAARHGLLASARQPGVCHYGTKLACCCYGNRRNSKGV 60
DB 1 MFLPWSLALPLLSSWAGGFGNNAARHGLLASARQPGVCHYGTKLACCCYGNRRNSKGV 60
QY 61 CEATCEPGCKFGCEGVGNKRCFPYGTGKTCSDQVNECGMKPRPCQHRVNTHGSYKFCF 120
DB 61 CEATCEPGCKFGCEGVGNKRCFPYGTGKTCSDQVNECGMKPRPCQHRVNTHGSYKFCF 120
QY 121 LSGHMLMPDATCVNSRTCAMINCYSCDETEEGPQCLCPSSGLRLAPNGRDCLDIDECAS 180
DB 121 LSGHMLMPDATCVNSRTCAMINCYSCDETEEGPQCLCPSSGLRLAPNGRDCLDIDECAS 180
QY 181 GKVICPNRRVNTFGSYCKCHIGFELQYISGRYDCIDINECTMDSHTCSHANCNTQ 240
DB 181 GKVICPNRRVNTFGSYCKCHIGFELQYISGRYDCIDINECTMDSHTCSHANCNTQ 240
QY 241 GSFCKCKQGYKGNGLRCSAIPENSVEVLRAPGTIKDRIKLLAHKNSMKKKAKIKNT 300
DB 241 GSFCKCKQGYKGNGLRCSAIPENSVEVLRAPGTIKDRIKLLAHKNSMKKKAKIKNT 300
QY 301 PEPTPTPKVNLQPNFYEEIVSRGNSHGKKGNEEKGLEDEKREKALKNDIEER 360
DB 301 PEPTPTPKVNLQPNFYEEIVSRGNSHGKKGNEEKGLEDEKREKALKNDIEER 360
QY 361 SLRGDVFPPKVNAGFGLLVORKALTSKLEHKLNI SVDCSFNHCICDWKQDREDDFD 420
DB 361 SLRGDVFPPKVNAGFGLLVORKALTSKLEHKLNI SVDCSFNHCICDWKQDREDDFD 420
QY 421 WNPADRDNAIGFYMAVPALAGHKD IGRLLKLLPDLPQPSNFCLLPDYRLAGDKVGLRV 480
DB 421 WNPADRDNAIGFYMAVPALAGHKD IGRLLKLLPDLPQPSNFCLLPDYRLAGDKVGLRV 480
QY 481 FVKNNSNALAWKTTSDDEKWKTKIQLYOGTDATKSI IPEARGKGTGEIADVGLLV 540
DB 481 FVKNNSNALAWKTTSDDEKWKTKIQLYOGTDATKSI IPEARGKGTGEIADVGLLV 540
QY 541 SGLCPDLSLSDDD 553
DB 541 SGLCPDLSLSDDD 553

Q9NY67 PRELIMINARY; PRT; 554 AA.
AC Q9NY67;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN W80.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Teratocarcinoma, and Neuron;
RA Franco B.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Teratocarcinoma, and Neuron;
RX MEDLINE=20241927; PubMed=10777661;
RA Buchner G., Orfanelli U., Quaderi N., Bassi M.T., Andolfi G.;
RT "Identification of a new EGF-repeat-containing gene from human Xp22.
RT Acandidate for developmental disorders.";
RL Genomics 65:16-23(2000).
CC -!- SIMILARITY: CONTAINS 1 MAM DOMAIN.
DR EMBL; AJ245671; CAB92132.1; -.
DR HSP; P00736; IAPQ.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR InterPro; IPR000152; Asx_hydroxyl_S.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR000998; MAM_domain.
DR Pfam; PF00008; EGF; 4.
DR Pfam; PF00629; MAM; 1.
DR SMART; SM00179; EGF_CA; 3.
DR PROSITE; PS00010; ASX_HYDROXYL; 3.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 2.
DR PROSITE; PS01187; EGF_CA; 3.
DR PROSITE; PS00060; MAM_2; 1.
KW Hypothetical protein; EGF-like domain; Glycoprotein.
SQ SEQUENCE 554 AA; 61388 MW; D519238F2A604101 CRC64;

Query Match 99.7%; Score 3049.5; DB 4; Length 554;
Best Local Similarity 99.8%; Pred. No. 2.5e-232;
Matches 553; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 1 MFLPWSLALPLLSSWAGGFGNNAARHGLLASARQPGVCHYGTKLACCCYGNRRNSKGV 60
DB 1 MFLPWSLALPLLSSWAGGFGNNAARHGLLASARQPGVCHYGTKLACCCYGNRRNSKGV 60
QY 61 CEATCEPGCKFGCEGVGNKRCFPYGTGKTCSDQVNECGMKPRPCQHRVNTHGSYKFCF 120
DB 61 CEATCEPGCKFGCEGVGNKRCFPYGTGKTCSDQVNECGMKPRPCQHRVNTHGSYKFCF 120
QY 121 LSGHMLMPDATCVNSRTCAMINCYSCDETEEGPQCLCPSSGLRLAPNGRDCLDIDECAS 180
DB 121 LSGHMLMPDATCVNSRTCAMINCYSCDETEEGPQCLCPSSGLRLAPNGRDCLDIDECAS 180
QY 181 GKVICPNRRVNTFGSYCKCHIGFELQYISGRYDCIDINECTMDSHTCSHANCNTQ 240
DB 181 GKVICPNRRVNTFGSYCKCHIGFELQYISGRYDCIDINECTMDSHTCSHANCNTQ 240
QY 241 GSFCKCKQGYKGNGLRCSAIPENSVEVLRAPGTIKDRIKLLAHKNSMKKKAKIKNT 300
DB 241 GSFCKCKQGYKGNGLRCSAIPENSVEVLRAPGTIKDRIKLLAHKNSMKKKAKIKNT 300
QY 301 PEPTPTPKVNLQPNFYEEIVSRGNSHGKKGNEEKGLEDEKREKALKNDIEER 360
DB 301 PEPTPTPKVNLQPNFYEEIVSRGNSHGKKGNEEKGLEDEKREKALKNDIEER 360

	QY	361	SURGDVFPKVNAGEFGILLIVORKALTSKLEHK-DLNI SVDCSFNHGICDWMQREDDF	419
	Db	361	SURGVFFFKVNAGEFGLIIVORKALTSKLEHKADLINI SVDCSFNHGI CDWKQREDDF	420
	QY	420	DWNPADRONALGFYNWAVPALAGHKKDIGRLKLLLPDLQPQSNCFLIFDYRLAGDKVGKLR	479
	Db	421	DWNPADRONAIGFYWAVPALAGHKKDIGRLKLLLPDLQPQSNCFLIFDYRLAGDKVGKLR	480
	QY	480	VVFVKNSNALAWEKTTSEDEKWTKGKIQLGYGTDTATKSIIFEARHGKGKTGEIANDGVLL	539
	Db	481	VVFVKNSNALAWEKTTSEDEKWTKGKIQLGYGTDTATKSIIFEARHGKGKTGEIANDGVLL	540
	QY	540	VSLGCPDSLLSVDD	553
	Db	541	VSLGCPDSLLSVDD	554
<hr/>				
	RESULT 5			
	Q9UFK6			
	ID	Q9UK6	PRELIMINARY;	PRT; 558 AA.
	AC	Q9UK6;		
	DT	01-MAY-2000	(TrEMBLrel. 13, Created)	
	DT	01-MAY-2000	(TrEMBLrel. 13, Last sequence update)	
	DT	01-OCT-2003	(TrEMBLrel. 25, Last annotation update)	
	DE	Hypothetical protein (Fragment).		
	DN	DKFPZP64P2063.		
	OS	Homo sapiens (Human).		
	OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
	OX	NCBI_TaxID=9606;		
	RN	[1]		
	RC	SEQUENCE FROM N.A.		
	RC	TTSUVE=Brain;		
	RA	Duesterhoeft A., Lauber J., Mewes H.W., Gassenhuber J., Wiemann S.;		
	RL	Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.		
	CC	-!- SIMILARITY: CONTAINS 1 MAM DOMAIN.		
	DR	EMBL; ALI117610; CAB56014.1; -.		
	DR	PIR; T17324; T17324.		
	DR	HSP; P00736; IAPQ.		
	DR	Gene; HGNC:3235; EGFL6.		
	DR	GO; GO:0005615; C:extracellular space; TAS.		
	DR	GO; GO:0005178; F:integrin binding; TAS.		
	DR	GO; GO:0007049; P:cell cycle; TAS.		
	DR	InterPro; IPRO00152; Asx hydroxyl_S.		
	DR	InterPro; IPRO01881; EGF Ca.		
	DR	InterPro; IPRO06209; EGF like.		
	DR	InterPro; IPRO00998; MAM_domain.		
	DR	pfam; PF00008; EGF; 4.		
	DR	pfam; PF00629; MAM; 1.		
	DR	SMART; SM00179; EGF_CA; 3.		
	DR	PROSITE; PS00010; ASX HYDROXYL; 3.		
	DR	PROSITE; PS00022; EGF 1; 1.		
	DR	PROSITE; PS01186; EGF 2; 2.		
	DR	PROSITE; PS01187; EGF CA; 2.		
	DR	PROSITE; PS50060; MAM 2; 1.		
	KW	Hypothetical protein_EGF-like domain; Glycoprotein.		
	FT	NON_TER		
	SQ	SEQUENCE	558 AA; 61828 MW; AA38D7DCE402BFA3 CRC64;	
<hr/>				
	Query Match		98.6%; Score 3016; DB 4; Length 558;	
	Best Local Similarity		98.9%; Pred. No. 1.1e-229;	
	Matches 547; Conservative		0; Mismatches 6; Indels 0; Gaps 0;	
<hr/>				
	QY	1	MLPWLSIALPLLSSWAGFGFNAAARHHGLLASAROPGVCHYGTKLACCYGWRNRSGV	60
	Db	6	MLPWLSIALPLLSSWAGFGFNAAARHHGLLASAROPGVCHYGTKLACCYGWRNRSGV	65
	QY	61	CEATCEPGCKFCGCVGNKRCFCGYTCKTCSDQNWEGMKPRPCQHRCVNTHGSVKFC	120
	Db	66	CEATCEPGCKFCGCVGNKRCCLPYTGKTCSDQNWEGMKPRPCQHRCVNTHGSVKFC	125
	QY	121	LSGHMLMPDATCVNSRTCAMINCOYSCEDTTEGPQCCLCPSSGLRAPNGRDCLDIDECAS	180

```
Qy 89 KTCSDVNECKMPPCOHRCVNTGSHGKFCCLSGHMLPMDATCVNSRTCAINCOYSCE 148
Db 10 ETCSDVNECKMPPCOHRCVNTGSHGKFCCLSGHMLPMDATCVNSRTCAINCOYSCE 69
Qy 149 DTEGPGQCLPSSGLRLAPNGRDCIDIDEACAGKVICPYNRRCVNTFGSYCKCHIGFEL 208
Db 70 DTEGPGQCLPSSGLRLAPNGRDCIDIDEACAGKVICPYNRRCVNTFGSYCKCHIGFEL 129
Qy 209 QYISGRYDCIDINECTMDSHCHSHANCFNTOGQFKCKCKGYKGNGLRCSAIPENSVE 268
Db 130 QYISGRYDCIDINECTMDSHCHSHANCFNTOGQFKCKCKGYKGNGLRCSAIPENSVE 189
Qy 269 VLAPGTTIKRIKLLAHKNSMKKAKIKNTPTPTPTPKVNIOPENYIEIVSRGNS 328
Db 190 VLAPGTTIKRIKLLAHKNSMKKAKIKNTPTPTPTPKVNIOPENYIEIVSRGNS 249
Qy 329 HGGKKGNEEKKEGLEDEKREKALKNDIEBRSRGDVFFPKVNEAGEFGLLVQRKALT 388
Db 250 HGGKKGNEEKKEGLEDEKREKALKNDIEBRSRGDVFFPKVNEAGEFGLLVQRKALT 309
Qy 389 SKLBHKOLNISVDCSFNHI CDWKQDREDDFDWNPADRDNAIGFYMAVPALAGHKQIGR 448
Db 310 SKLBHKOLNISVDCSFNHI CDWKQDREDDFDWNPADRDNAIGFYMAVPALAGHKQIGR 369
Qy 449 LKLLLPDLQPOSNFCLLPDYRLAGDKVGLRVFKNSNNALAWKTTSEDEKWKTKIQL 508
Db 370 LKLLLPDLQPOSNFCLLPDYRLAGDKVGLRVFKNSNNALAWKTTSEDEKWKTKIQL 429
Qy 509 YQGTDTATSIIFAEARGKGTGEIADVGLLVSLGLCPDLSLSD 553
Db 430 YQGTDTATSIIFAEARGKGTGEIADVGLLVSLGLCPDLSLSD 474

RESULT 7
Q9JUZ5 PRELIMINARY; PRT; 550 AA.
AC Q9JUZ5;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN EGFL6 OR W80.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Franco B.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20241927; PubMed=10777661;
RA Buchner G., Orfanelli U., Quaderi N., Bassi M.T., Andolfi G.;
RT "Identification of a new EGF-repeat-containing gene from human Xp22:
RT Candidates for developmental disorders.";
RL Genomics 65:16-23(2000).
CC -!- SIMILARITY: CONTAINS 1 MAM DOMAIN.
DR EMBL; AJ245672; CAB92138.1; -.
DR HSSP; P35555; 1EMN.
DR MGI; MGI:1858599; Egfl6.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR InterPro; IPR000152; Asx_hydroxyl_S.
DR InterPro; IPR001881; EGF Ca.
DR InterPro; IPR006209; EGF like.
DR Pfam; PF00008; EGF; 4.
DR Pfam; PF00629; MAM; 1.
DR SMART; SM00179; EGF_CA; 3.
DR PROSITE; PS00010; ASX_HYDROXYL; 3.
DR PROSITE; PS00022; EGF_1; 1.
```

```
DR PROSITE; PS01186; EGF_2; 2.
DR PROSITE; PS01187; EGF_CA; 3.
DR PROSITE; PS00060; MAM_2; 1.
KW Hypothetical protein; EGF-like domain; Glycoprotein.
SQ SEQUENCE 550 AA; 61520 MW; DEF936325C9F31B3 CRC64;

Query Match 78.5%; Score 2401.5; DB 11; Length 550;
Best Local Similarity 77.9%; Pred. No. 3.7e-181;
Matches 43; Conservative 48; Mismatches 69; Indels 5; Gaps 4;

Qy 1 MFLPWSLALPLLSSVAGFGNAASARHHGLLASARPGVCHYGTKLACCYGWRNKGV 60
Db 1 MQPWLGLALPLLPLPWTGGVG--TSPWDYGLSALAHQPGVQYGTQMACCYGWRNKGV 58
Qy 61 CEATCEPGCFEGCVGNKCRCPGYTGKTSQDVNECKMPPCOHRCVNTGSHGKFC 120
Db 59 CEATCEPGCFEGCVGNKCRCPGYTGKTCQDVNECKMPPCOHRCVNTGSHGKFC 118
Qy 121 LSGHMLPMDATCVNSRTCAINCOYSCEDETEGQCLPSSGLRLAPNGRDCIDIDEACAS 180
Db 119 LSGHMLPMDATCVNSRTCAINCOYSCEDETEGQCLPSSGLRLAPNGRDCIDIDEACAS 178
Qy 181 GKVTCPVNRRCVNTFGSYCKCHIGFELQYISGRYDCIDINECTMDSHCHSHANCFNTQ 240
Db 179 SKAVCPENRRRCVNTFGSYCKCHIGFELQYISGRYDCIDINECTMDSHCHSHANCFNTR 238
Qy 241 GSPKCKCKQYKGNGLRCSAIPENSVEKVLRAPTTKDKRIKLLAHKNSMKKAKIKNT 300
Db 239 GSPKCKCKQYKGNGLRCSAIPENSVEKVLRAPTTKDKRIKLLAHKNSMKKAKIKNT 298
Qy 301 PEPTPTPTPKVNIOPENYIEIVSRGNSHGKKEBEKKEGLEDEKREKALKNDIE-E 359
Db 299 PRASTRVPKVN- PYSSSEGVSRGNYDGEQKKKEEKEERLEEEK-GEKTLRNEVEQE 356
Qy 360 RSLRGDVFFPKVNEAGEFGLLVQRKALTSLBHKOLNISVDCSFNHI CDWKQDREDD 419
Db 357 RSLRGDVFFPKVNEAGEFGLLVQRKELNSKLKHKOLNISVDCSFNHI CDWKQDREDD 416
Qy 420 DWNPADRDNAIGFYMAVPALAGHKQIGRLKLLPDLQPOSNFCLLPDYRLAGDKVGLR 479
Db 417 DWNPADRDNDVGYMAVPALAGHKQIGRLKLLPDLQPOSNFCLLPDYRLAGDKVGLR 476
Qy 480 VFVKNSSNNALAWKTTSEDEKWKTKIQLYQGTDTATSIIFAEARGKGTGEIADVGLLV 539
Db 477 VFVKNSSNNALAWKTTSEDEKWKTKIQLYQGTDTATSIIFAEARGKGTGEIADVGLLV 536
Qy 540 VSLGLCPDLSLSD 552
Db 537 VSLGLCPDLSLSD 549

RESULT 8
Q8AVH7 PRELIMINARY; PRT; 544 AA.
ID Q8AVH7
AC Q8AVH7;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Similar to EGF-like-domain, multiple 6.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE=Embryo;
RA Klein S., Strausberg R.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC042275; AAH42275.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR InterPro; IPR000152; Asx_hydroxyl_S.
```

```
DR InterPro: IPR001881; EGF_CA.
DR InterPro: IPR006209; EGF_like.
DR InterPro: IPR006210; IEFG.
DR InterPro: IPR00998; MAM_domain.
DR Pfam: PF00008; EGF_3.
DR Pfam: PF00629; MAM; 1.
DR PRINTS: PR00020; MAMDOMAIN.
DR SMART: SM00181; EGF; 5.
DR SMART: SM00179; EGF_CA; 3.
DR SMART: SM00137; MAM; 1.
DR PROSITE: PS00010; ASX HYDROXYL; 3.
DR PROSITE: PS00022; EGF_1; 1.
DR PROSITE: PS01186; EGF_2; 3.
DR PROSITE: PS01187; EGF_CA; 3.
DR PROSITE: PS00600; MAM 2; 1.
SQ SEQUENCE 544 AA; 61048 MW; 24865A62DEAB7B5 CRC64;

Query Match 53.3%; Score 1630.5; DB 13; Length 544;
Best Local Similarity 56.1%; Pred. No. 2.8e-120;
Matches 294; Conservative 77; Mismatches 142; Indels 11; Gaps 5;

QY 24 ASARHGLLASARQGVCHYGTKLACCYGWRNRSGVCEATCEPCKGECVGNPKRCF 83
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
30 SSRSHRLITSPSTTGVCRYGKAECCYGKRNKRGQCEAVCEQGCKGECVGNPKRCF 89
QY 84 PGTGKTOSDVNECGMKPRPCQHRVCVTHGSKYKFCLSCHMLMPDATCVNSRTCAMINC 143
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
90 PGTGKNCNQLNECLGKPRPCQHRVCVTHGSKYKFCLSCHMLMPDGSCNSRTCAMANC 149
QY 144 QVSCDETEGPOCLPSSGLRLAPNGRDCLDIDECASGKVICPNRRCVNTFGSYCKCH 203
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
150 QVCGQVKGDIRCLCPSSGLQLGPDGRTCIDIDECAGKASCPINRRCVNTFGSYCKCQ 209
QY 204 IGFELOYISGRYDCIDINECTWDSHTCSHHANCFNTQGSFKCKQGYKGNGLRCSAIP 263
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
210 IGYELKYVNGRYDCIDINECLLTHKCSINADCLNTQGSFKCKCKGFKGNQCECAVEN 269
QY 264 NSVKEVLRAPGTIKRIKLLAHKNSMCKKAKIKNVTPPTTPKVNLPFNVEEIVS 323
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
270 KPVKSPKFGSGVQKAIKLLAHKNSLNRYNDIKNVIPETFTIPPPKRLQDFVEDGVY 329
QY 324 RGNSHGGKKEEKKMEGLEDKEEKKALNDIDERSLRGDFVFPKVNAGEFGLILVQ 383
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
330 ICGNUNDEEGIEEIEEELDEEENVE---EKLRLGDFARQVKRAA---VLSS 382
QY 384 RKALFSKLEHKDLNTSVDCSFNHGICQKQDREDDPDWNPARDNAIGFYMAVPALAGHK 443
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
383 QPISNTDPLVKDEVLVDCRFQDGTCEWKQDSKDDFDWKHAERHNGYINVPASTSQK 442
QY 444 KDIGRLKLLPDLQPOSNFCLLFDYRLAGDKYGLRVFVKNSNNAALAWETTSDEKWK 503
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
443 KGIGRLKQLTKI--YKYLCLMFTYRLAGERVKURVYIDENINPI--WEETKNRDEGWT 499
QY 504 GKILYQ-ETDATKSTIIFAEKGTGKTBIAVDGLLVSLGCPD 546
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
500 AKIEQESSTRKSSSTTFEAVRGKDEAGIMALDNYFLSSGPGCD 543

RESULT 9
Q8BPM8 PRELIMINARY; PRT; 327 AA.
AC Q8BPM8;
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE EGF-like domain (Fragment).
GN EGF16.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RC STRAIN=C57BL/6J; TISSUE=Eye;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
DR EMBL; AK053738; BACJ5499.1; -.
DR MGD; MGI:1858599; Eglf6.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR InterPro: IPR00152; Asx hydroxyl_s.
DR InterPro: IPR001881; EGF_CA.
DR InterPro: IPR006209; EGF_like.
DR InterPro: IPR008210; IEFG.
DR Pfam; PF00008; EGF; 4.
DR SMART; SM00181; EGF; 5.
DR SMART; SM00179; EGF_CA; 3.
DR PROSITE; PS00010; ASX HYDROXYL; 3.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 2.
DR PROSITE; PS01187; EGF_CA; 3.
FT NON_TER 327 327
SQ SEQUENCE 327 AA; 36063 MW; 69751FD528CE756D CRC64;

Query Match 48.6%; Score 1486.5; DB 11; Length 327;
Best Local Similarity 77.9%; Pred. No. 3.4e-109;
Matches 257; Conservative 25; Mismatches 45; Indels 3; Gaps 2;

QY 1 MPLPWSLALPLLSSVAGGFGNAASARHHGLLASARQGVCHYGTKLACCYGWRNRSGV 60
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1 MQPWLGLALPLLPLPWTGGVG--TSPWDYGLSALAHQPGVQCYGTQWACCYGWRNRSGV 58
QY 61 CEATCEPCKGECVGNPKRCFPYGTGKTCSDVNECGMKPRPCQHRVCVTHGSKYKFC 120
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
59 CEATCEPCKGECVGNPKRCFPYGTGKTCQDVNECGMKPRPCQHRVCVTHGSKYKFC 118
QY 121 LSGHMLMPDATCVNSRTCAMINCOYSCEDTEGPOCLPSSGLRLAPNGRDCLDIDECAS 180
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
119 LSGHMLPDATCVNSRTCARLNCQYGCDETEGPRCVCPSSGLRLGPNRGVCLDIDECAS 178
QY 181 GKVICPNRRCVNTFGSYCKCHIGFELOYISGRYDCIDINECTWDSHTCSHHANCFNTQ 240
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
179 SKAVCPNRRVCVNTFGSYCKCHIGFELOYISGRYDCIDINECTWDSHTCSHHANCFNTQ 238
QY 241 GSFCKCKQGYKGNGLRCSAIPENSVEVLRAPGTIKRIKLLAHKNSMCKKAKIKNV 300
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
239 GSFCKCKQGYKGNGLRCSAIPENSVEVLRAPGTIKRIKLLAHKNSMCKKAKIKNV 298
QY 301 PEPTPTTPKVNLPFNVEEIVSRGNSHG 330
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
299 PPASTRVPKVNL-PYSSEEGVSRGRNYDG 327

RESULT 10
Q91V88 PRELIMINARY; PRT; 561 AA.
AC Q91V88;
DT 01-DEC-2001 (TRENBLrel. 19, Created)
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE POEM (NEPHROECTIN short isoform).
GN NPNT OR POEM OR NEPH1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RX MEDLINE=2155216; PubMed=11546798;
RA Morimura N., Tezuka Y., Watanabe N., Yasuda M., Miyatani S.,
RA Hozumi N., Tezuka K.;
RP "Molecular cloning of POEM. A novel adhesion molecule that interacts
```

RT with alpha2beta1 integrin.";
 RL J. Biol. Chem. 276:42172-42181 (2001).
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NIH Swiss; TISSUE=Kidney;
 RX MEDLINE=21363579; PubMed=11470831;
 RA Brandenberger R., Schmidt A., Linton J., Wang D., Backus C., Denda S.,
 RA Muller U., Reichardt L.F.,
 RT "Identification and characterization of a novel extracellular matrix
 RT protein nephronectin that is associated with integrin alpha2beta1 in
 RT the embryonic kidney.";
 RL J. Cell Biol. 154:447-458 (2001).
 [3]
 RN SEQUENCE FROM N.A.
 RP STRAIN=C57BL/6J; TISSUE=Pancreas;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs.";
 RL Nature 420:563-573 (2002).
 CC -1- SIMILARITY: CONTAINS 1 MAM DOMAIN.
 DR EMBL; AB059656; BAB69692.1; -;
 DR EMBL; AY035898; AAK96010.1; -;
 DR EMBL; AK050484; BAC34283.1; -;
 DR MGD; MGI:2148811; Npnt.
 DR GO; GO:0005578; C:extracellular matrix; IDA.
 DR GO; GO:0030022; F:adhesive extracellular matrix constituent a. . .; IDA.
 DR GO; GO:0005178; F:integrin binding; IDA.
 DR GO; GO:0007160; P:cell-matrix adhesion; IDA.
 DR InterPro; IPR000152; Asx_hydroxyl_S.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR000998; MAM_domain.
 DR Pfam; PF00008; EGF; 4.
 DR Pfam; PF00629; MAM; 1.
 DR SMART; SM00179; EGF_CA; 2.
 DR PROSITE; PS00010; ASX_HYDROXYL; 3.
 DR PROSITE; PS00022; EGF_1; 1.
 DR PROSITE; PS01186; EGF_2; 3.
 DR PROSITE; PS01187; EGF_CA; 3.
 DR PROSITE; PS00060; MAM_2; 1.
 DR EGF-like domain; Glycoprotein.
 SK SEQUENCE 561 AA; 61490 MW; 69E7ACAAB0EB3F506 CRC64;
 QY Query Match 35.8%; Score 1095; DB 11; Length 561;
 Db Best Local Similarity 38.0%; Pred. No. 5, 7e-78;
 Matches 210; Conservative 92; Mismatches 180; Indels 70; Gaps 11;
 39 GVCHYGTGLACCYGWRNRSKGVCEATCPGKFGCEVGNKCRCPFGYTGKTCSDVNEC 98
 34 GLCRYGRIDCCWGRQSWGQCQPVCPQCKHGEVGNKCKCHPGAGTKCNQDLNEC 93
 99 GMKPRPCQHRVNTGSKYKCLSHMLMPDATCVNSTCAMINCOYSCDETEGPOCLC 158
 94 GLKPRPCQHRVNTGSKYKCLNGYMLPDPGSCSSALSCSNAMCQYCDVWKGQVRCQ 153
 159 PSSGLRLAPNGDCLIDICACGKVICPNRRVNTFGSYCKCHIGFELYISRYDCI 218
 154 PPSGLQLAPDGTQVDIDECATGRVSCFRPQCQVNTFGSYCKCHTGTDFDLYIGKQYCH 213
 219 DINECTMDSHTCSSHANCFNTQSPKCKQKQYKGNGLRCSAIPENSVK-----EVLRAP 273
 214 DIDECSLGHQHCSSVARYCNIHSGYKQCRDGYEGDGLNCVYIPKVMIEPSGPIHMPERN 273
 274 GTIK-----DRIKLLAHKNSMKKAK-IKNVTPETPTTPK 310
 274 GTISGDGGHANRIPDAGSTRWPLNTPYIPVITNRPSTKPTRTPNPTPTPTPTPTPTPTPT 333
 311 VNLQPFNVEEIVSRGCGHGGKKGNEKKEGLEDEKREKAL-----KN 355
 334 LPTEP-----RTPLPTPPTPPT 374

QY 356 DIEERSLRGVDFFPKVNEAGEFGLLVORLKALTSKLEHKDLNLSV-DCSPNHHGICDKQD 414
 Db 375 QTPDQKPRGVDFFPRQPTNDLFFEIFEIRGVSADDEEVKDDPGILIHSCNFDHGLCGWIRE 434
 QY 415 REDDFDWNPADRDNAIGFYMAVPALAGHKDIGRLKLLLPDLPQPSNFCLLFDYRLAGDK 474
 Db 435 KDSDLHWETA-RDPAGGYLTVSAKAPGGAARLVRLGHLHMSGDLCLSFHKKVTLGLH 493
 QY 475 VGKLRVFK--NSNNALAWBKTTSEDEKWKTKIQLYQGTDAKSIIFEAERGKGTGEI 532
 Db 494 SGTQLQVFRKHGTHGAALWGRNGHG--WRQQTITL-RGAD-VKSVIFPKGKRRTGTEI 549
 QY 533 AVDGVLVSLGLC 544
 Db 550 GLDDVSLKRGRC 561
 RESULT 11
 Q91ZD3 PRELIMINARY; PRT; 578 AA.
 AC Q91ZD3;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Nephronectin long isoform.
 GN NPNT OR NEPH1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 OX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NIH Swiss; TISSUE=Kidney;
 RX MEDLINE=21363579; PubMed=11470831;
 RA Brandenberger R., Schmidt A., Linton J., Wang D., Backus C., Denda S.,
 RA Muller U., Reichardt L.F.,
 RT "Identification and characterization of a novel extracellular matrix
 RT protein nephronectin that is associated with integrin alpha2beta1 in
 RT the embryonic kidney.";
 RL J. Cell Biol. 154:447-458 (2001).
 CC -1- SIMILARITY: CONTAINS 1 MAM DOMAIN.
 DR EMBL; AY035899; AAK96011.1; -;
 DR MGD; MGI:2148811; Npnt.
 DR GO; GO:0005578; C:extracellular matrix; IDA.
 DR GO; GO:0030022; F:adhesive extracellular matrix constituent a. . .; IDA.
 DR GO; GO:0005178; F:integrin binding; IDA.
 DR GO; GO:0007160; P:cell-matrix adhesion; IDA.
 DR InterPro; IPR000152; Asx_hydroxyl_S.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR000998; MAM_domain.
 DR Pfam; PF00008; EGF; 4.
 DR Pfam; PF00629; MAM; 1.
 DR SMART; SM00179; EGF_CA; 2.
 DR PROSITE; PS00010; ASX_HYDROXYL; 3.
 DR PROSITE; PS00022; EGF_1; 1.
 DR PROSITE; PS01186; EGF_2; 3.
 DR PROSITE; PS01187; EGF_CA; 3.
 DR PROSITE; PS00060; MAM_2; 1.
 DR EGF-like domain; Glycoprotein.
 SQ SEQUENCE 578 AA; 63606 MW; 4757DE54CC47DA4C CRC64;
 Query Match 35.3%; Score 1081.5; DB 11; Length 578;
 Best Local Similarity 37.1%; Pred. No. 6, 9e-77;
 Matches 211; Conservative 92; Mismatches 179; Indels 87; Gaps 12;
 39 GVCHYGTGLACCYGWRNRSKGVCE-----ATCEPCKFGCEVGNKCR 81
 34 GLCRYGRIDCCWGRQSWGQCQPFYVLRQRRLARIFCQLKAVCQPCQCKHGEVGNKCK 93
 82 CFPQYTKTCSQDVNECGMKPRPCQHRVNTGSKYKCLSHMLMPDATCVNSTCAMI 141
 94 CHPGFAGTKCNQDLNEGLKPRCPCKHGMNTFGSKYKCYCLNGYMLLPDGGSCSSALSCSA 153

Similar to nephronectin (Fragment).

DE Mus musculus (Mouse).

OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

NCBI_TaxID=10090;

RN [1]_SEQUENCE FROM N.A.

RP STRAIN=C57BL/6J; TISSUE=Breast tumor;

RA Strausberg K.; Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.

RL EMBL; BC046642; AAH46642.1; -.

DR GO; GO:0016020; C:membrane; IEA.

DR GO; GO:0005509; F:calcium ion binding; IEA.

DR InterPro; IPR000152; Asx_hydroxyl_S.

DR InterPro; IPR001881; EGF_CA.

DR InterPro; IPR006209; EGF_like.

DR InterPro; IPR006210; IEGF.

DR InterPro; IPR000998; MAM_domain.

DR Pfam; PF00008; EGF; 1.

DR Pfam; PF00629; MAM; 1.

DR SMART; SM00181; EGF; 1.

DR SMART; SM00179; EGF_CA; 1.

DR SMART; SM00137; MAM; 1.

DR PROSITE; PS00010; ASX_HYDROXYL; 1.

DR PROSITE; PS01186; EGF_2; 1.

DR PROSITE; PS01187; EGF_CA; 1.

DR PROSITE; PSS0060; MAM_2; 1.

FT NON_TER 1

SQ SEQUENCE 372 AA; 40909 MW; F654EDC6AAC34517 CRC64;

Query Match 15.4%; Score 472; DB 11; Length 372;

Best Local Similarity 29.0%; Pred. No. 5.3e-29;

Matches 115; Conservative 61; Mismatches 150; Indels 70; Gaps 11.

QY 195 FGSYYCKCHIGFLEQYISGYDCIDINECTMDSHTCSHHANCFNTQGSFKCKKQGYKGN 254

DB 1 FGSYICKCHTGFDLMYIGGKYQCHDIDECISLGQHQCSSYARCVNIHSGYKQCRDGYEGD 60

QY 255 GLRCSAIPENVK-----EVLRAPGTIK-----DRIKKLAAHK 287

DB 61 GLNCVYIPKVMIPSPGSIHPERNGTISKDGGHANIPDAGSTRWPLKTPYIPVITNR 120

QY 288 NSMKKKAK-IKNVTPETPTKPNVLQPNFYEBIVSRGGNSHGKKGNEEKMEGLEDE 346

DB 121 PTKSPTRPTNPPTPQPTPPPPPLPTP-----RTTLPPTPERP 161

QY 347 KREKAL-----KNDIERSLRGVFPKVNKEAFGLLVQKALTSKL 391

DB 162 STRPTTIAPATSTTRVITVDNRITQDPKPRGVDFIPRQPTNDLFEIERGVSADEE 221

QY 392 EHKDLNTSV-DCSNHNGICDWQBEDDFDWNADRDNAIGFYMAVPALAGHKDGLRK 450

DB 222 VKDDPGIILHSCNFDHGLCGWIREKDSLHWETA-RDPAGGQYLTVSAKAPGGKAARLV 280

QY 451 LLLPDLQPQSNFCLLDYRLAGDKVGLRVFVK--NSNNALAWEKTTSEDEKWTGKIOL 508

DB 281 LRLGHLMHSGDLCLSFRRKVTGLSHSGTLOVFRKHGTGHAALWGRNGCHG--WRQTQITL 338

QY 509 YQGTDATKSIIFEAERGKGTGETIAVDGVLVSLGLC 544

DB 339 -RGAD-VKSVIFKGEKRRGHTGEIGLDDVSLKRGRC 372

RESULT 15

Q9WUH9 PRELIMINARY; PRT: 2906 AA.

ID Q9WUH9

AC Q9WUH9

DT 01-NOV-1999 (TRENBLrel. 12, Created)

DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)

DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)

DE Fibrillin-2.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99350231; PubMed=10419698;
RA Yang Q., Ota K., Tian Y., Kumar A., Wada J., Kashiwara N., Wallner E.,
RA Kanwar Y.S.;
RT "Cloning of rat fibrillin-2 cDNA and its role in branching
RT morphogenesis of embryonic lung";
RL Dev. Biol. 212:229-242(1999).
DR EMBL; AF135060; AAD34439.1; -.
DR HSSP; P35555; 1EMN.
DR GO; GO:000509; F:calcium ion binding; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR002086; Aldehyde dehydr.
DR InterPro; IPR000152; Asx_hydroxyl_S.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR002212; Fibril-assoc.
DR Pfam; PF00008; EGF; 44.
DR Pfam; PF00683; TE; 9.
DR SMART; SM00179; EGF_CA; 42.
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; 43.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS01186; EGF_2; 36.
DR PROSITE; PS01187; EGF_CA; 43.
KW EGF-like domain.
SQ SEQUENCE 2906 AA; 313371 MW; 9EE64E727044EF58 CRC64;

Query Match 13.3%; Score 408.5; DB 11; Length 2906;
Best Local Similarity 29.3%; Pred. No. 7.8e-23;
Matches 98; Conservative 32; Mismatches 92; Indels 113; Gaps 13;

QY 38 PGVCHYGT-----KLACCVG-----RRNSKGVCEATCEPG-CKFGEV---GPNKC 80
DQ 1073 PGMCTYGRNRTIGSFKRCNSGFALDMEERNCTDIDECRISPDLCGNGICVNTGSPFEC 1132
QY 81 RCFFGYTG-----KTCSDVNECGMKPRPQ-HRCVNTGSGYKFCFLSGHMLP----- 128
DQ 1133 ECFEGYESGFMKMKC-WDIDECERNPILCRGTCVNTGSEFQDCPLGHELSPREDCI 1191
QY 129 -----DATCVNRTCAMI-----NCQYSCED 149
DQ 1192 DINECSLDNLCRNGKCVNMIGTYQSCNPGYQATPDQGCSDIDECIMMGCDTQCTN 1251
QY 150 TEEGQCILCPSSGLRPNRCLDIDECASGVIC-----PYNRECV----- 192
DQ 1252 SEGSYECSC-SEGYALMPDGRSCADIDECENPDICDGGQCTNIPGEYRCLCYDGFMSM 1310
QY 193 -----NTFGSYCKCHIGFELQYISGRYDCIDINECTMDSH 228
DQ 1311 DMKTCIDVNECLNPNICMFGECENTKGSFICHQQLGYSVK--KGATGCTDVDECEIGA 1368
QY 229 TCSHANCNTQGSFKCKQKQYKNGLRCSAIP 263
DQ 1369 NCDMHASCLNVFGSKCSREGVNGVNGIKCIDILDE 1403

Search completed: May 21, 2004, 12:33:15
Job time : 72.0986 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 21, 2004, 12:26:11 ; Search time 2.72007 Seconds

(without alignments)
1558.126 Million cell updates/sec

Title: US-09-981-649a-24_COPY_412_426

Perfect score: 89

Sequence: 1 KQDREDDFDWNPADR 15

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq_29Jan04.*

1: geneseqp1980s.*

2: geneseqp1990s.*

3: geneseqp2000s.*

4: geneseqp2001s.*

5: geneseqp2002s.*

6: geneseqp2003as.*

7: geneseqp2003bs.*

8: geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	89	100.0	502	AAO15367	AAO15367 Human EGF
2	89	100.0	502	AAE26499	AAE26499 Human epi
3	89	100.0	502	ABG72941	ABG72941 Novel hum
4	89	100.0	502	ABU62264	ABU62264 Epidermal
5	89	100.0	537	AAI18110	AAI18110 Protein e
6	89	100.0	537	AAO15360	AAO15360 Human EGF
7	89	100.0	537	AAE26498	AAE26498 Human epi
8	89	100.0	537	ABG72934	ABG72934 Novel hum
9	89	100.0	537	ABU62257	ABU62257 Epidermal
10	89	100.0	551	ABG72294	ABG72294 Rat prote
11	89	100.0	553	AAI18108	AAI18108 Protein e
12	89	100.0	553	AAO1423	AAO1423 Human TAN
13	89	100.0	553	AAI18110	AAI18110 Human pol
14	89	100.0	553	AAO15361	AAO15361 Human EGF
15	89	100.0	553	AAO15368	AAO15368 Human EGF
16	89	100.0	553	AAE26506	AAE26506 Human epi
17	89	100.0	553	AAE26500	AAE26500 Human epi
18	89	100.0	553	ABJ05586	ABJ05586 Breast ca
19	89	100.0	553	ABG72942	ABG72942 Novel hum
20	89	100.0	553	ABG72935	ABG72935 Novel hum
21	89	100.0	553	ABR48234	ABR48234 Human bla
22	89	100.0	553	ABU56725	ABU56725 Lung canc
23	89	100.0	553	ABU62265	ABU62265 Epidermal
24	89	100.0	553	ABU62258	ABU62258 Epidermal
25	89	100.0	553	ADB80482	ADB80482 Ovarian c

26	89	100.0	554	4	AB27224	Human EXM
27	89	100.0	554	4	AA39156	Human pol
28	89	100.0	554	5	AAO15371	Human EGF
29	89	100.0	554	5	AAO15370	Human EGF
30	89	100.0	554	6	ABG72945	Novel hum
31	89	100.0	554	6	ABG72944	Novel hum
32	89	100.0	554	6	ABU62267	Novel epi
33	89	100.0	554	6	ABU62268	Novel epi
34	89	100.0	559	5	AAO15369	Human EGF
35	89	100.0	559	6	ABG72943	Novel hum
36	89	100.0	559	6	ABU62266	Novel epi
37	89	100.0	573	4	AAO40942	Human pol
38	84	94.4	14	6	ABG72946	Novel hum
39	73.5	82.6	15	6	ABU62269	Epidermal
40	53	59.6	339	6	ABR41667	Human DIT
41	53	59.6	1718	5	AAE26420	Human tra
42	50	56.2	262	6	ABU11894	Human ABC
43	50	56.2	310	4	AAU33035	Novel hum
44	47	52.8	292	4	ABB66682	Drosophil
45	46.5	52.2	875	7	ADE57383	Rat Prote

ALIGNMENTS

RESULT 1

AAO15367
ID AAO15367 standard; protein; 502 AA.

XX AAO15367;

XX 19-SEP-2002 (first entry)

XX Human EGF motif-containing protein, SEQ ID No 18.

Human; epidermal growth factor motif; EGF motif; EGF μ 6;
epithelial tissue growth; tissue repair; tissue regeneration;
corneal transplant healing; skin graft; wound healing; cancer; leukaemia;
nervous system disorder; infection; autoimmune disorder; inflammation;
multiple sclerosis; anaemia; periodontal disease; haemophilia;
fertility enhancement.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Misc-difference 501 /note= "Encoded by GAN"

FT Misc-difference 502 /note= "Encoded by NNC"

FT WO200230977-A2.

XX 18-APR-2002.

XX 15-OCT-2001; 2001WO-US032257.

XX 13-OCT-2000; 2000US-00687860.

XX (HYSE-) HYSEQ INC.

XX Asundi V, Ford JE, Drmanac RT, Liu C, Yamasaki V, Yeung G;

XX Tang TV, Zhang J, Zhou P, Zhou H;

XX WPI; 2002-426270/45.

XX N-ESDB; AAL43889.

Novel isolated epidermal growth factor motif polypeptide, termed EGF μ 6,
for treating cancer, nervous system disorders, immune deficiencies,
autoimmune disorders, coagulation disorders and inflammatory conditions.

XX Disclosure; Page 162-163; 183pp; English.

XX The invention comprises the amino acid and coding sequences of human

CC epidermal growth factor (EGF) motif-containing proteins (EGFL6 proteins).
 CC The DNA and protein sequences of the invention are useful for inhibiting
 CC the proliferation of cells expressing an EGFL6 protein. The DNA and
 CC protein sequences of the invention are useful for stimulating epithelial
 CC tissue growth, for tissue repair and regeneration, corneal transplant
 CC healing, skin graft production and wound healing. The DNA and protein
 CC sequences are useful for treating cancer, leukaemia, nervous system
 CC disorders, infection, autoimmune disorders (e.g. multiple sclerosis),
 CC anaemia, periodontal diseases, haemophilia, inflammatory conditions, and
 CC for effecting bodily characteristics and fertility of male or female
 CC subjects. The present amino acid sequence represents a human EGF motif-
 CC containing protein
 CC
 XX Sequence 502 AA;
 SQ

Query Match 100.0%; Score 89; DB 5; Length 502;
 Best Local Similarity 100.0%; Pred. No. 1.2e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KQDREDDFDWNPADR 15
 |||||
 DB 361 KQDREDDFDWNPADR 375

RESULT 2
 AAE26499
 ID AAE26499 standard; protein; 502 AA.
 XX
 AC AAE26499;
 XX
 DT 13-DEC-2002 (first entry)
 DE Human epidermal growth factor (EGF)-repeat containing protein #2.
 DE
 XX Human; antibody; epidermal growth factor; EGF repeat; brain tumour;
 KW nervous disorder; ulcer; leukaemia.
 XX
 OS Homo sapiens.
 XX

Key Location/Qualifiers
 FH Key
 FT Misc-difference 84
 FT /note= "Amino acid Xaa is present at this position in the
 FT sequence shown in column 59-62 of the specification"
 FT /label= Unknown
 FT /note= "Xaa can be any amino acid"
 FT Misc-difference 502
 FT /label= Unknown
 FT /note= "Xaa can be any amino acid"
 FT
 FT
 XX US6392019-B1.
 XX
 XX 21-MAY-2002.
 XX
 XX 28-JUL-1999; 99US-00363316.
 XX
 XX 22-NOV-1997; 97US-00968800.
 XX 12-FEB-1999; 99US-00249697.
 XX
 XX (FORD/) FORD J.
 XX (YEUN/) YEUNG G.
 XX
 XX Ford J, Yeung G;
 XX
 XX WPI; 2002-424836/45.
 XX N-PSDB; AAD44331.
 XX
 XX Novel antibody specific for an epidermal growth factor repeat-containing
 PT polypeptide, useful for the diagnosis of brain tumors, ulcers, leukemias,
 PT and nervous disorders.
 XX
 XX Disclosure; Col 81-84; 92pp; English.

CC The invention relates to an antibody specific for a 537 residue epidermal
 CC growth factor (EGF) repeat-containing polypeptide sequence. The invention
 CC is used for detecting the presence of EGF repeat containing polypeptides
 CC in a sample, in the diagnosis of brain tumours, nervous disorders,
 CC ulcers, and leukemias. The present sequence is human EGF-repeat
 CC containing protein
 XX
 SQ Sequence 502 AA;
 Query Match 100.0%; Score 89; DB 5; Length 502;
 Best Local Similarity 100.0%; Pred. No. 1.2e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KQDREDDFDWNPADR 15
 |||||
 DB 361 KQDREDDFDWNPADR 375

RESULT 3
 ABG72941
 ID ABG72941 standard; protein; 502 AA.
 XX
 AC ABG72941;
 XX
 DT 02-APR-2003 (first entry)
 DE Novel human EGF-motif containing protein fragment #3.
 DE
 XX EGF; epidermal growth factor; cancer; lung cancer; brain cancer;
 KW prostate cancer; breast cancer; skin cancer; lymphoma cancer;
 KW sarcoma cancer; colon cancer; tumorigenicity; tumour site reduction;
 KW cell proliferation inhibition; vaccine; antisense gene therapy; human.
 XX
 OS Homo sapiens.
 XX

Key Location/Qualifiers
 FH Key
 FT Misc-difference 501
 FT /label= OTHER
 FT /note= "Any amino acid"
 FT Misc-difference 502
 FT /label= OTHER
 FT /note= "Any amino acid"
 FT
 XX US2002132250-A1.
 XX
 XX 19-SEP-2002.
 XX
 XX 15-OCT-2001; 2001US-00981649.
 XX
 XX 28-JUL-1999; 99US-00363316.
 XX 13-OCT-2000; 2000US-00687860.
 XX
 XX (FORD/) FORD J E.
 XX (YEUN/) YEUNG G.
 XX (ZHOU/) ZHOU H.
 XX
 XX Ford JE, Yeung G, Zhou H;
 XX
 XX WPI; 2003-174078/17.
 XX
 XX Detecting cancerous cells expressing polynucleotides/polypeptides in
 PT samples, by contacting samples with labeled polynucleotides complementary
 PT to polynucleotide or an antibody against the polypeptide and detecting
 PT complex formed.
 XX
 XX Disclosure; Page 53-54; 78pp; English.
 XX
 XX The invention describes a method of detecting a cancerous cell expressing
 CC a polynucleotide (I) or a polypeptide (II) in a biological sample,
 CC involving contacting the sample with a labelled polynucleotide
 CC complementary to (I) or an antibody or its fragment that specifically
 CC binds to (II), for a period sufficient to form a complex and detecting
 CC the complex, so that if a complex is detected, the cell is detected. The

CC identify agents which bind to them and the nucleotide sequences can be
 CC used as probes for in situ hybridisation. The polypeptides and their
 CC polynucleotides can also be used for other therapeutic, diagnostic and
 CC research utilities
 XX
 SQ Sequence 537 AA;
 Query Match 100.0%; Score 89; DB 2; Length 537;
 Best Local Similarity 100.0%; Pred. No. 1.2e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 KQREDDFDWNPADR 15
 Db 361 KQREDDFDWNPADR 375
 RESULT 6
 AAO15360
 ID AAO15360 standard; protein; 537 AA.
 XX
 AC AAO15360;
 XX
 DT 19-SEP-2002 (first entry)
 XX
 DE Human EGF motif-containing protein, SEQ ID No 4.
 XX
 KW Human; epidermal growth factor motif; EGF motif; EGFL6;
 KW epithelial tissue growth; tissue repair; tissue regeneration;
 KW corneal transplant healing; skin graft; wound healing; cancer; leukaemia;
 KW nervous system disorder; infection; autoimmune disorder; inflammation;
 KW multiple sclerosis; anaemia; periodontal disease; haemophilia;
 KW fertility enhancement.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 501 /note= "Encoded by GAN"
 FT Misc-difference 502 /note= "Encoded by NNC"
 FT Misc-difference 503 /note= "Xaa is encoded by a stop codon (TGA)"
 XX
 PN WO200230977-A2.
 XX
 PD 18-APR-2002.
 XX
 PF 15-OCT-2001; 2001WO-US032257.
 XX
 PR 13-OCT-2000; 2000US-00687860.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Asundi V, Ford JB, Drmanac RT, Liu C, Yamasaki V, Yeung G;
 PI Tang TY, Zhang J, Zhou P, Zhou H;
 XX
 DR WPI; 2002-426270/45.
 DR N-PSDB; AAL43899.
 XX
 CC Novel isolated epidermal growth factor motif polypeptide, termed EGFL6,
 CC for treating cancer, nervous system disorders, immune deficiencies,
 CC autoimmune disorders, coagulation disorders and inflammatory conditions.
 XX
 PS Example 1; Fig 3; 183pp; English.
 XX
 CC The invention comprises the amino acid and coding sequences of human
 CC epidermal growth factor (EGF) motif-containing proteins (EGFL6 proteins).
 CC The DNA and protein sequences of the invention are useful for inhibiting
 CC the proliferation of cells expressing an EGFL6 protein. The DNA and
 CC protein sequences of the invention are useful for stimulating epithelial
 CC tissue growth, for tissue repair and regeneration, corneal transplant
 CC healing, skin graft production and wound healing. The DNA and protein
 CC sequences are useful for treating cancer, leukaemia, nervous system

CC disorders, infection, autoimmune disorders (e.g. multiple sclerosis),
 CC anaemia, periodontal diseases, haemophilia, inflammatory conditions, and
 CC for effecting bodily characteristics and fertility of male or female
 CC subjects. The present amino acid sequence represents a human EGF motif-
 CC containing protein
 XX
 SQ Sequence 537 AA;
 Query Match 100.0%; Score 89; DB 5; Length 537;
 Best Local Similarity 100.0%; Pred. No. 1.2e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 KQREDDFDWNPADR 15
 Db 361 KQREDDFDWNPADR 375
 RESULT 7
 AAE26498
 ID AAE26498 standard; protein; 537 AA.
 XX
 AC AAE26498;
 XX
 DT 13-DEC-2002 (first entry)
 XX
 DE Human epidermal growth factor (EGF)-repeat containing protein #3.
 XX
 KW Human; antibody; epidermal growth factor; EGF repeat; brain tumour;
 KW nervous disorder; ulcer; leukaemia.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 503 /label= Unknown
 FT /note= "Xaa can be any amino acid"
 XX
 PN US6392019-B1.
 XX
 PD 21-MAY-2002.
 XX
 PF 28-JUL-1999; 99US-00363316.
 XX
 PR 22-NOV-1997; 97US-00968800.
 PR 12-FEB-1999; 99US-00249697.
 XX
 PA (FORD/) FORD J.
 PA (YEUN/) YEUNG G.
 XX
 PI Ford J, Yeung G;
 XX
 DR WPI; 2002-424836/45.
 XX
 CC Novel antibody specific for an epidermal growth factor repeat-containing
 CC polypeptide, useful for the diagnosis of brain tumors, ulcers, leukemias,
 CC and nervous disorders.
 XX
 PS Claim 1; Fig 3; 92pp; English.
 XX
 CC The invention relates to an antibody specific for a 537 residue epidermal
 CC growth factor (EGF) repeat-containing polypeptide sequence. The invention
 CC is used for detecting the presence of EGF repeat containing polypeptides
 CC in a sample, in the diagnosis of brain tumors, nervous disorders,
 CC ulcers, and leukemias. The present sequence is human EGF-repeat
 CC containing protein
 XX
 SQ Sequence 537 AA;
 Query Match 100.0%; Score 89; DB 5; Length 537;
 Best Local Similarity 100.0%; Pred. No. 1.2e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 KQREDDFDWNPADR 15

Best Local Similarity 100.0%; Pred. No. 1.2e-05; Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KQREDDFDWNPADR 15
 DB 361 KQREDDFDWNPADR 375

RESULT 9
 ABU62257
 ID ABU62257 standard; protein; 537 AA.
 XX
 AC ABU62257;
 XX
 DT 01-SEP-2003 (first entry)
 XX
 DE Epidermal growth factor motif protein EGFL6 fragment #2.
 XX
 KW Human; epidermal growth factor motif protein; EGFL6; cytostatic;
 KW neuroprotective; antibacterial; antiparasitic; antilepemic;
 KW antiinfectility; EGF-Agonist; EGF-Antagonist; cell growth; cancer;
 KW neurodegenerative disorder; leukaemia; brain tumour; lung tumour;
 KW breast tumour; gastrointestinal tumour; skin tumour; prostate tumour;
 KW carcinoma; parasite; biorhythm; fertility; metabolism; catabolism;
 KW anabolism.
 XX
 OS Homo sapiens.
 XX
 PH Key Location/Qualifiers
 FT Misc-difference 503 /label= OTHER
 FT /note= "OTHER= any amino acid"
 FT
 XX US2003036508-A1.
 XX
 XX 20-FEB-2003.
 XX
 XX 17-APR-2002; 2002US-00124986.
 XX
 XX 22-NOV-1997; 97US-00968800.
 XX 12-FEB-1999; 99US-00249697.
 XX 28-JUL-1999; 99US-00363316.
 XX 13-OCT-2000; 2000US-00687860.
 XX 15-OCT-2001; 2001US-00981649.
 XX
 XX (FORD/) FORD J.
 XX (YEUN/) YEUNG G.
 XX (ZHOU/) ZHOU H.
 XX
 XX Ford J, Yeung G, Zhou H;
 XX
 XX WPI; 2003-492123/46.
 XX N-PSDB; ACD25930.
 XX
 XX Stimulating cell growth by contacting the cell with an EGFL6 polypeptide,
 XX useful for the diagnosis and treatment of cancers and neurodegenerative
 XX disorders.
 XX
 XX Example 3; Fig 3; 86pp; English.
 XX
 XX The invention describes a method of stimulating cell growth comprising
 XX contacting the cell with an EGFL6 polypeptide having at least 90 %
 XX sequence identity to a 553 amino acid sequence (S1), given in the
 XX specification, or its variant and/or fragment lacking a C-terminal
 XX portion of the EGFL6 polypeptide. The methods and compositions of the
 XX present invention are useful for the diagnosis and treatment of cancers
 XX and neurodegenerative disorders by stimulating cell growth. The cancers
 XX include leukaemia, brain, lung, breast, gastrointestinal, skin and
 XX prostate tumours and carcinomas. They can also be used in inhibiting the
 XX growth of infectious agents and parasites, effecting bodily
 XX characteristics and biorhythms, effecting fertility, metabolism
 XX catabolism and anabolism of fats, vitamins, proteins and minerals, and
 XX effecting behavioural characteristics. This is the amino acid sequence of

Db 361 KQREDDFDWNPADR 375

RESULT 8
 ABG72934
 ID ABG72934 standard; protein; 537 AA.
 XX
 AC ABG72934;
 XX
 DT 02-APR-2003 (first entry)
 XX
 DE Novel human EGF-motif containing protein fragment #2.
 XX
 KW EGF; epidermal growth factor; cancer; lung cancer; brain cancer;
 KW prostate cancer; breast cancer; skin cancer; lymphoma cancer;
 KW sarcoma cancer; colon cancer; tumorigenicity; tumour site reduction;
 KW cell proliferation inhibition; vaccine; antisense gene therapy; human.
 XX
 OS Homo sapiens.
 XX
 PH Key Location/Qualifiers
 FT Misc-difference 501 /note= "Encoded by GAN"
 FT Misc-difference 502 /note= "Encoded by NNC"
 FT Misc-difference 503 /note= "Encoded by TGA"
 FT
 XX US2002132250-A1.
 XX
 XX 19-SEP-2002.
 XX
 XX 15-OCT-2001; 2001US-00981649.
 XX
 XX 28-JUL-1999; 99US-00363316.
 XX 13-OCT-2000; 2000US-00687860.
 XX
 XX (FORD/) FORD J E.
 XX (YEUN/) YEUNG G.
 XX (ZHOU/) ZHOU H.
 XX
 XX Ford JE, Yeung G, Zhou H;
 XX
 XX WPI; 2003-174078/17.
 XX N-PSDB; ABX14767.
 XX
 XX Detecting cancerous cells expressing polynucleotides/polypeptides in
 XX samples, by contacting samples with labeled polynucleotides complementary
 XX to polynucleotide or an antibody against the polypeptide and detecting
 XX complex formed.
 XX
 XX Example 1; Fig 3; 78pp; English.
 XX
 XX The invention describes a method of detecting a cancerous cell expressing
 XX a polynucleotide (I) or a polypeptide (II) in a biological sample,
 XX involving contacting the sample with a labelled polynucleotide
 XX complementary to (I) or an antibody or its fragment that specifically
 XX binds to (II), for a period sufficient to form a complex and detecting
 XX the complex, so that if a complex is detected, the cell is detected. The
 XX method is useful for detecting cancerous cell in a biological sample such
 XX as tissue, cell, blood, serum, lymphatic fluid, urine, and cerebrospinal
 XX fluid. The cancerous cell is from lung, brain, prostate, breast, skin,
 XX lymphoma, sarcoma and colon. Preferably the cancer cell is A549 cell, MCF
 XX -7 cell or SK-N-Mc cell. PC1 and PC2 are useful for inhibiting EGFL6
 XX activity, inhibiting tumorigenicity, reducing tumour sites and inhibiting
 XX proliferation of a cancer cell. This is the amino acid sequence of a
 XX novel human EGF (epidermal growth factor) motif containing protein
 XX fragment

Sequence 537 AA:
 SQ Query Match 100.0%; Score 89; DB 6; Length 537;

CC a novel human epidermal growth factor motif protein EGFL6 fragment
XX
SQ Sequence 537 AA;

Query Match 100.0%; Score 89; DB 6; Length 537;
Best Local Similarity 100.0%; Pred. No. 1.2e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KQREDDEDWNPADR 15
|||||
Db 361 KQREDDEDWNPADR 375

RESULT 10
ABB72294
ID ABB72294 standard; protein; 551 AA.

XX ABB72294;
XX
DT 04-APR-2002 (first entry)
DE Rat protein isolated from skin cells SEQ ID NO: 506.

XX Human; rat; mouse; skin cell; skin wound; cancer; growth defect;
KW developmental defect; inflammatory disease; dermatological; vulvar;
KW immunomodulator; anti-inflammatory; cytostatic; neuroprotective.

XX Rattus sp.
XX WO200190357-A1.
PN
XX
XX 29-NOV-2001.

XX 24-MAY-2001; 2001WO-NZ000099.
XX
XX 24-MAY-2000; 2000US-0206650P.
PR 25-JUL-2000; 2000US-0221232P.
XX (GENE-) GENESIS RES & DEV CORP LTD.

XX Watson JD, Strachan L, Sleeman M, Onrust R, Murison JG;
PI Kumble XD;
PI
XX WPI; 2002-122020/16.

XX New polynucleotides and polypeptides encoded by the polynucleotides
PT isolated from skin cells, useful for treating skin wounds, cancers,
PT growth and developmental defects, inflammatory diseases, or for
PT modulating immune responses.

XX Claim 4; Page 311-312; 466pp; English.

XX The present invention provides the protein and coding sequences of cDNAs
CC isolated from human, murine and rat skin cell libraries. The sequences
CC can be used in the development of therapeutic agents useful in the
CC treatment of skin diseases, including skin wounds, cancer, growth
CC defects, developmental defects and inflammatory diseases. The proteins
CC have important roles in the induction of hair growth, cell proliferation
CC and cell-cell interaction, in maintaining tissue integrity, in wound
CC healing and in modulating immune responses. The present sequence is a
CC polypeptide of the invention

XX Sequence 551 AA;

Query Match 100.0%; Score 89; DB 5; Length 551;
Best Local Similarity 100.0%; Pred. No. 1.3e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KQREDDEDWNPADR 15
|||||
Db 410 KQREDDEDWNPADR 424

RESULT 11
AAV18108
ID AAV18108 standard; protein; 553 AA.
XX
XX AAV18108;
AC
XX
DT 10-AUG-1999 (first entry)
DE Protein encoded by cDNA insert of clone pEGFR-HY2.

XX Epidermal growth factor; EGF repeat domain; haematopoiesis regulator;
KW tissue growth activity; activin; inhibitor; chemotaxis; chemokinesis;
KW haemostasis; thrombolysis; anti-inflammatory; leukaemia; anaemia;
KW immune disorder; immune deficiency; nervous system disorder; therapy.

XX Synthetic.
XX WO9927096-A1.
PN
XX
XX 03-JUN-1999.

XX 23-NOV-1998; 98WO-US024524.

XX 22-NOV-1997; 97US-00968800.

XX (HYSE-) HYSEQ INC.

PI Drmanac RT, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
PI Leshkowitz D, Kita D, Ford J;
XX
XX WPI; 1999-370904/31.

DR N-PSDB; AAX79501.

XX New polypeptide with epidermal growth factor repeat domains.

XX Claim 8; Fig 5; 96pp; English.

XX This sequence represents a polypeptide of the invention, which has
CC similarity to epidermal growth factor (EGF) repeat domains. The
CC polypeptides and their compositions may have haematopoiesis regulating,
CC tissue growth, activin/inhibin, chemotactic/chemokinetic, haemostatic,
CC thrombolytic, receptor/ligand and anti-inflammatory activities. They may
CC be used to treat leukaemias, anaemias, immune disorders and deficiencies
CC and nervous system disorders. They can be used in screening assays to
CC identify agents which bind to them and the nucleotide sequences can be
CC used as probes for in situ hybridisation. The polypeptides and their
CC polynucleotides can also be used for other therapeutic, diagnostic and
CC research utilities

XX Sequence 553 AA;

Query Match 100.0%; Score 89; DB 2; Length 553;
Best Local Similarity 100.0%; Pred. No. 1.3e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KQREDDEDWNPADR 15
|||||
Db 412 KQREDDEDWNPADR 426

RESULT 12
AAB01423
ID AAB01423 standard; protein; 553 AA.

XX AAB01423;

XX 20-OCT-2000 (first entry)

XX Human TANGO 212.

XX TANGO; 128; 140; 197; 212; 224; 239; modulating agent; asthma;
KW graft versus-host diseases; rheumatoid arthritis; psoriasis;
KW inflammatory bowel disease; septic shock; ulcerative colitis;

KW Crohn's disease; chronic myelogenous leukemia; cancer; liver disease;
KW Hodgkin's disease; osteoarthritis; Lyme's disease; cachexia;
KW autoimmune disease; myasthenia gravis; autoimmune diabetes;
KW systemic lupus erythematosus; transgenic animal; diagnosis; prognosis;
KW prophylactic; therapeutic; human.
XX Homo sapiens.
XX WO200039284-A1.
XX 06-JUL-2000.
XX 23-DEC-1999; 99WO-US031025.
XX 30-DEC-1998; 98US-00223546.
XX (MILL-) MILLENNIUM PHARM INC.
XX Holtzman DA;
XX WPI; 2000-465743/40.
XX N-PSDB; AAA47456.
XX Novel nucleic acid sequences encoding TANGO-128, 140, 197, 212, 213, 224
XX and 239 polypeptides useful for the treatment of asthma, rheumatoid
XX arthritis, psoriasis and autoimmune diseases.
XX Claim 8; Fig 5; 209pp; English.
XX Nucleic acids encoding TANGO polypeptides are useful as modulating agents
XX for regulating cellular processes like asthma, graft versus-host
XX diseases, rheumatoid arthritis, psoriasis, inflammatory bowel disease,
XX septic shock, ulcerative colitis, Crohn's disease, chronic myelogenous
XX leukemia, cancer, liver disease, Hodgkin's disease, osteoarthritis,
XX Lyme's disease, cachexia and autoimmune diseases e.g. myasthenia gravis,
XX autoimmune diabetes and systemic lupus erythematosus. The nucleic acids
XX are also useful for producing transgenic animals and the TANGO
XX polypeptides themselves. Partial TANGO-128, 140, 197, 212, 213, 224, 239
XX sequences are useful in forensic biology, for diagnostic assays,
XX prognostic assays, pharmacogenomics and for monitoring clinical trials.
XX TANGO polypeptides are suitable for both prophylactic and therapeutic
XX methods for treating a subject at risk of a disorder or having a disorder
XX associated with aberrant TANGO expression. A wide range of cellular
XX disorders can be treated
XX SQ Sequence 553 AA;
Query Match 100.0%; Score 89; DB 3; Length 553;
Best Local Similarity 100.0%; Pred. No. 1.3e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KQREDDFDWNPADR 15
Db 412 KQREDDFDWNPADR 426
RESULT 13
AAM93622
ID AAM93622 standard; protein; 553 AA.
XX AAM93622;
XX AAM93622;
XX 06-NOV-2001 (first entry)
XX Human polypeptide, SEQ ID NO: 3456.
XX Human; full length cDNA; cDNA synthesis; oligo-capping.
XX Homo sapiens.
XX EP1130094-A2.
XX PF 15-OCT-2001; 2001WO-US032257.
XX PD 05-SEP-2001.

XX 07-JUL-2000; 2000EP-00114089.
PF
XX 08-JUL-1999; 99JP-00194486.
XX 11-JAN-2000; 2000JP-00118774.
PR 02-MAY-2000; 2000JP-00183765.
XX (HELI-) HELIX RES INST.
XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX WPI; 2001-524255/58.
DR N-PSDB; AAK94555.
XX 830 Primers useful for synthesizing full length cDNA clones and their use
XX in genetic manipulation.
XX Claim 8; SEQ ID NO 3456; 1380pp + Sequence Listing; English.
XX The invention relates to primers for synthesising full length cDNA
XX clones. 830 cDNA molecules encoding a human protein have been isolated
XX and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have
XX been determined. Primers for synthesising the full length cDNA are useful
XX for clarifying the function of the protein encoded by the cDNA. The full
XX length clones were obtained by construction of full length enriched cDNA
XX libraries that were synthesised by the oligo-capping method. The primers
XX enable the production of the full length cDNA easily without any special
XX methods. The present sequence is a polypeptide encoded by a full length
XX human cDNA of the invention. Note: The sequence data for this patent did
XX not form part of the printed specification, but was obtained in CD-ROM
XX format directly from EPO
XX SQ Sequence 553 AA;
Query Match 100.0%; Score 89; DB 4; Length 553;
Best Local Similarity 100.0%; Pred. No. 1.3e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KQREDDFDWNPADR 15
Db 412 KQREDDFDWNPADR 426
RESULT 14
AAO15361
ID AAO15361 standard; protein; 553 AA.
XX AAO15361;
XX 19-SEP-2002 (first entry)
XX Human EGF motif-containing protein, SEQ ID No 6.
XX Human; epidermal growth factor motif; EGF motif; EGFL6;
XX epithelial tissue growth; tissue repair; tissue regeneration;
XX corneal transplant healing; skin graft; wound healing; cancer; leukaemia;
XX nervous system disorder; infection; autoimmune disorder; inflammation;
XX multiple sclerosis; anaemia; periodontal disease; haemophilia;
XX fertility enhancement.
XX Homo sapiens.
XX Key Location/Qualifiers
XX Key Mismatch 357 /note= "Encoded by WTA"
XX FT WO200230977-A2.
XX PN 18-APR-2002.
XX PD 15-OCT-2001; 2001WO-US032257.
XX PF
XX

```

PR 13-OCT-2000; 2000US-00687860.
XX (HYSE-) HYSEQ INC.
XX Asundi V, Ford JE, Drmanac RT, Liu C, Yamasaki V, Yeung G;
XX Tang TY, Zhang J, Zhou P, Zhou H;
XX WPI: 2002-426270/45.
XX N-PSDB; AAL43890.
XX Novel isolated epidermal growth factor motif polypeptide, termed EGFL6,
XX for treating cancer, nervous system disorders, immune deficiencies,
XX autoimmune disorders, coagulation disorders and inflammatory conditions.
XX Example 3; Fig 5; 183pp; English.
XX The invention comprises the amino acid and coding sequences of human
XX epidermal growth factor (EGF) motif-containing proteins (EGFL6 proteins).
XX The DNA and protein sequences of the invention are useful for inhibiting
XX the proliferation of cells expressing an EGFL6 protein. The DNA and
XX protein sequences of the invention are useful for stimulating epithelial
XX tissue growth, for tissue repair and regeneration, corneal transplant
XX healing, skin graft production and wound healing. The DNA and protein
XX sequences are useful for treating cancer, leukaemia, nervous system
XX disorders, infection, autoimmune disorders (e.g. multiple sclerosis),
XX anaemia, periodontal diseases, haemophilia, inflammatory conditions, and
XX for effecting bodily characteristics and fertility of male or female
XX subjects. The present amino acid sequence represents a human EGF motif-
XX containing protein
XX Sequence 553 AA;
SQ
Query Match 100.0%; Score 89; DB 5; Length 553;
Best Local Similarity 100.0%; Pred. No. 1.3e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KQDREDDFDWNPADR 15
DB 412 KQDREDDFDWNPADR 426

RESULT 15
AAO15368
XX ID AAO15368 standard; protein; 553 AA.
XX AC AAO15368;
XX 19-SEP-2002 (first entry)
XX Human EGF motif-containing protein, SEQ ID No 24.
XX Human; epidermal growth factor motif; EGF motif; EGFL6;
XX epithelial tissue growth; tissue repair; tissue regeneration;
XX corneal transplant healing; skin graft; wound healing; cancer; leukaemia;
XX nervous system disorder; infection; autoimmune disorder; inflammation;
XX multiple sclerosis; anaemia; periodontal disease; haemophilia;
XX fertility enhancement.
XX Homo sapiens.
XX WO200230977-A2.
XX 18-APR-2002.
XX 15-OCT-2001; 2001WO-US032257.
XX 13-OCT-2000; 2000US-00687860.
XX (HYSE-) HYSEQ INC.
XX Asundi V, Ford JE, Drmanac RT, Liu C, Yamasaki V, Yeung G;
XX Tang TY, Zhang J, Zhou P, Zhou H;

```

```

DR WPI: 2002-426270/45.
XX N-PSDB; AAL43901.
XX Novel isolated epidermal growth factor motif polypeptide, termed EGFL6,
XX for treating cancer, nervous system disorders, immune deficiencies,
XX autoimmune disorders, coagulation disorders and inflammatory conditions.
XX Claim 28; Page 167-169; 183pp; English.
XX The invention comprises the amino acid and coding sequences of human
XX epidermal growth factor (EGF) motif-containing proteins (EGFL6 proteins).
XX The DNA and protein sequences of the invention are useful for inhibiting
XX the proliferation of cells expressing an EGFL6 protein. The DNA and
XX protein sequences of the invention are useful for stimulating epithelial
XX tissue growth, for tissue repair and regeneration, corneal transplant
XX healing, skin graft production and wound healing. The DNA and protein
XX sequences are useful for treating cancer, leukaemia, nervous system
XX disorders, infection, autoimmune disorders (e.g. multiple sclerosis),
XX anaemia, periodontal diseases, haemophilia, inflammatory conditions, and
XX for effecting bodily characteristics and fertility of male or female
XX subjects. The present amino acid sequence represents a human EGF motif-
XX containing protein
XX Sequence 553 AA;
SQ
Query Match 100.0%; Score 89; DB 5; Length 553;
Best Local Similarity 100.0%; Pred. No. 1.3e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KQDREDDFDWNPADR 15
DB 412 KQDREDDFDWNPADR 426

Search completed: May 21, 2004, 12:31:18
Job time : 3.72007 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 21, 2004, 12:29:27 ; Search time 0.871479 Seconds
(without alignments)
888.592 Million cell updates/sec

Title: US-09-981-649A-24_COPY_412_426

Perfect score: 89
Sequence: 1 KQREDDFDWNPADR 15

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*

1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PTCUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	89	100.0	502	4	US-09-363-316B-18
2	89	100.0	537	4	US-09-249-697A-4
3	89	100.0	537	4	US-09-363-316B-4
4	89	100.0	553	4	US-09-249-697A-6
5	89	100.0	553	4	US-09-249-697A-19
6	89	100.0	553	4	US-09-363-316B-6
7	89	100.0	553	4	US-09-363-316B-24
8	43	48.3	147	1	US-08-688-609-8
9	43	48.3	147	3	US-09-002-832-8
10	41	46.1	227	4	US-09-071-035-476
11	41	46.1	290	4	US-09-543-681A-8241
12	41	46.1	454	4	US-09-071-035-252
13	41	46.1	894	4	US-09-071-035-248
14	41	46.1	962	4	US-09-071-035-246
15	41	46.1	962	4	US-09-071-035-250
16	41	46.1	962	4	US-09-071-035-254
17	41	46.1	962	4	US-09-071-035-470
18	41	46.1	962	4	US-09-071-035-474
19	41	46.1	962	4	US-09-071-035-478
20	41	46.1	970	4	US-09-134-000C-5691
21	41	46.1	4655	3	US-08-652-877-86
22	40	44.9	232	4	US-09-489-039A-13759
23	40	44.9	943	3	US-08-476-515A-12
24	40	44.9	944	3	US-08-652-877-12
25	40	44.9	4654	3	US-08-476-515A-84
26	40	44.9	4655	3	US-08-652-877-84
27	40	44.9	4655	3	US-08-652-877-88

28 40 44.9 4655 3 US-08-652-877-90 Sequence 90, Appl
29 39.5 44.4 163 4 US-09-615-192A-306 Sequence 306, App
30 39 43.8 146 1 US-08-688-609-6 Sequence 6, Appli
31 39 43.8 146 3 US-09-002-832-6 Sequence 6, Appli
32 39 43.8 146 4 US-09-976-594-830 Sequence 830, App
33 39 43.8 148 1 US-08-688-609-3 Sequence 3, Appli
34 39 43.8 148 3 US-09-002-832-3 Sequence 3, Appli
35 39 43.8 457 3 US-08-821-818-1 Sequence 1, Appli
36 39 43.8 457 3 US-08-821-818-1 Sequence 5, Appli
37 39 43.8 457 4 US-09-052-753B-1 Sequence 1, Appli
38 39 43.8 457 4 US-09-052-753B-5 Sequence 5, Appli
39 39 43.8 531 2 US-09-933-750C-9 Sequence 9, Appli
40 39 43.8 531 3 US-09-234-613-9 Sequence 9, Appli
41 39 43.8 531 4 US-09-647-143-2 Sequence 2, Appli
42 39 43.8 571 4 US-09-328-352-7033 Sequence 7033, Ap
43 39 43.8 578 4 US-09-052-753B-7 Sequence 7, Appli
44 39 43.8 913 4 US-08-971-089-4 Sequence 4, Appli
45 38.5 43.3 648 4 US-09-437-687A-16 Sequence 16, Appli

ALIGNMENTS

RESULT 1
US-09-363-316B-18
; Sequence 18, Application US/09363316B
; Patent No. 6392019
; GENERAL INFORMATION:
; APPLICANT: Ford, John
; APPLICANT: Yeung, George
; TITLE OF INVENTION: EGF MOTIF PROTEIN MATERIALS AND METHODS
; FILE REFERENCE: 28110/35852
; CURRENT APPLICATION NUMBER: US/09/363,316B
; CURRENT FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: US 09/249,697
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 08/968,800
; PRIOR FILING DATE: 1997-11-22
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 18
; LENGTH: 502
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; LOCATION: (501-502)
; OTHER INFORMATION: Xaa = any amino acid
US-09-363-316B-18

Query Match 100.0%; Score 89; DB 4; Length 502;
Best Local Similarity 100.0%; Pred. No. 4.4e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KQREDDFDWNPADR 15
Db 361 KQREDDFDWNPADR 375

RESULT 2
US-09-249-697A-4
; Sequence 4, Application US/09249697A
; Patent No. 6392018
; GENERAL INFORMATION:
; APPLICANT: Ford, John
; APPLICANT: Yeung, George
; TITLE OF INVENTION: NOVEL EGF MOTIF PROTEIN OBTAINED FROM A CDNA LIBRARY OF FETAL
; FILE REFERENCE: 24011-727
; CURRENT APPLICATION NUMBER: US/09/249,697A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 08/968,800
; PRIOR FILING DATE: 1997-11-22

```
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 537
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(537)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-249-697A-4
```

```
Query Match 100.0%; Score 89; DB 4; Length 537;
Best Local Similarity 100.0%; Pred. No. 4.7e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 KQDREDDFDWNPADR 15
Db 361 KQDREDDFDWNPADR 375
```

```
RESULT 3
US-09-363-316B-4
; Sequence 4, Application US/09363316B
; Patent No. 6392019
; GENERAL INFORMATION:
; APPLICANT: Ford, John
; APPLICANT: Yeung, George
; TITLE OF INVENTION: EGF MOTIF PROTEIN MATERIALS AND METHODS
; FILE REFERENCE: 28110/35852
; CURRENT APPLICATION NUMBER: US/09/363,316B
; CURRENT FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: US 09/249,697
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 08/968,800
; PRIOR FILING DATE: 1997-11-22
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 537
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (503)
; OTHER INFORMATION: Xaa = any amino acid
US-09-363-316B-4
```

```
Query Match 100.0%; Score 89; DB 4; Length 537;
Best Local Similarity 100.0%; Pred. No. 4.7e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 KQDREDDFDWNPADR 15
Db 361 KQDREDDFDWNPADR 375
```

```
RESULT 4
US-09-249-697A-6
; Sequence 6, Application US/09249697A
; Patent No. 6392018
; GENERAL INFORMATION:
; APPLICANT: Ford, John
; APPLICANT: Yeung, George
; TITLE OF INVENTION: NOVEL EGF MOTIF PROTEIN OBTAINED FROM A CDNA LIBRARY OF FETAL
; TITLE OF INVENTION: LIVER SPLEEN
; FILE REFERENCE: 24011-727
; CURRENT APPLICATION NUMBER: US/09/249,697A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 08/968,800
; PRIOR FILING DATE: 1997-11-22
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 3.0
```

```
; SEQ ID NO 6
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(553)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-249-697A-6
```

```
Query Match 100.0%; Score 89; DB 4; Length 553;
Best Local Similarity 100.0%; Pred. No. 4.8e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 KQDREDDFDWNPADR 15
Db 412 KQDREDDFDWNPADR 426
```

```
RESULT 5
US-09-249-697A-19
; Sequence 19, Application US/09249697A
; Patent No. 6392018
; GENERAL INFORMATION:
; APPLICANT: Ford, John
; APPLICANT: Yeung, George
; TITLE OF INVENTION: NOVEL EGF MOTIF PROTEIN OBTAINED FROM A CDNA LIBRARY OF FETAL
; TITLE OF INVENTION: LIVER SPLEEN
; FILE REFERENCE: 24011-727
; CURRENT APPLICATION NUMBER: US/09/249,697A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 08/968,800
; PRIOR FILING DATE: 1997-11-22
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 19
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-249-697A-19
```

```
Query Match 100.0%; Score 89; DB 4; Length 553;
Best Local Similarity 100.0%; Pred. No. 4.8e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 KQDREDDFDWNPADR 15
Db 412 KQDREDDFDWNPADR 426
```

```
RESULT 6
US-09-363-316B-6
; Sequence 6, Application US/09363316B
; Patent No. 6392019
; GENERAL INFORMATION:
; APPLICANT: Ford, John
; APPLICANT: Yeung, George
; TITLE OF INVENTION: EGF MOTIF PROTEIN MATERIALS AND METHODS
; FILE REFERENCE: 28110/35852
; CURRENT APPLICATION NUMBER: US/09/363,316B
; CURRENT FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: US 09/249,697
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 08/968,800
; PRIOR FILING DATE: 1997-11-22
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
```


Fri May 21 12:56:10 2004

```
QY      4 REDDFDNPADR 15
      :|||||
Db      72 QEDDLFWPPDR 83

      TYPE: PRT
      ORGANISM: Proteus mirabilis
      US-09-543-681A-8241

      Query Match      46.1%; Score 41; DB 4; Length 290;
      Best Local Similarity 42.9%; Pred. No. 64;
      Matches 6; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

      QY      1 QDREDDFDNPNAD 14
      Db      236 KADKDEHFTWLPD 249

      RESULT 12
      US-09-071-035-252
      ; Sequence 252, Application US/09071035
      ; Patent No. 6448043
      ; GENERAL INFORMATION:
      ; APPLICANT: Gil H. Choi
      ; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
      ; NUMBER OF SEQUENCES: 496
      ; CORRESPONDENCE ADDRESS:
      ; ADDRESSEE: Human Genome Sciences, Inc.
      ; STREET: 9410 Key West Avenue
      ; CITY: Rockville
      ; STATE: Maryland
      ; COUNTRY: USA
      ; ZIP: 20850
      ; COMPUTER READABLE FORM:
      ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
      ; COMPUTER: HP Vectra 486/33
      ; OPERATING SYSTEM: MSDOS version 6.2
      ; SOFTWARE: ASCII Text
      ; CURRENT APPLICATION DATA:
      ; APPLICATION NUMBER: US/09/071,035
      ; FILING DATE:
      ; CLASSIFICATION:
      ; PRIOR APPLICATION DATA:
      ; APPLICATION NUMBER:
      ; FILING DATE:
      ; ATTORNEY/AGENT INFORMATION:
      ; NAME: A. Anders Brookes
      ; REGISTRATION NUMBER: 36,373
      ; REFERENCE/DOCKET NUMBER: PB369P2
      ; TELEPHONE: (301) 309-8504
      ; TELEFAX: (301) 309-8512
      ; INFORMATION FOR SEQ ID NO: 476:
      ; SEQUENCE CHARACTERISTICS:
      ; LENGTH: 227 amino acids
      ; TYPE: amino acid
      ; STRANDEDNESS: single
      ; TOPOLOGY: linear
      ; MOLECULE TYPE: protein
      ; US-09-071-035-476

      Query Match      46.1%; Score 41; DB 4; Length 227;
      Best Local Similarity 66.7%; Pred. No. 50;
      Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

      QY      3 DREDDFDN 11
      Db      46 NREDSFDWH 54

      RESULT 11
      US-09-543-681A-8241
      ; Sequence 8241, Application US/09543681A
      ; Patent No. 6605709
      ; GENERAL INFORMATION:
      ; APPLICANT: GARY BRETON
      ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
      ; FILE REFERENCE: 2709.1002-001
      ; CURRENT APPLICATION NUMBER: US/09/543,681A
      ; CURRENT FILING DATE: 2000-04-05
      ; PRIOR APPLICATION NUMBER: US 60/128,706
      ; PRIOR FILING DATE: 1999-04-09
      ; NUMBER OF SEQ ID NOS: 8344
      ; SEQ ID NO 8241
      ; LENGTH: 290

      TYPE: PRT
      ORGANISM: Proteus mirabilis
      US-09-543-681A-8241

      Query Match      46.1%; Score 41; DB 4; Length 290;
      Best Local Similarity 42.9%; Pred. No. 64;
      Matches 6; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

      QY      1 QDREDDFDNPNAD 14
      Db      236 KADKDEHFTWLPD 249

      RESULT 12
      US-09-071-035-252
      ; Sequence 252, Application US/09071035
      ; Patent No. 6448043
      ; GENERAL INFORMATION:
      ; APPLICANT: Gil H. Choi
      ; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
      ; NUMBER OF SEQUENCES: 496
      ; CORRESPONDENCE ADDRESS:
      ; ADDRESSEE: Human Genome Sciences, Inc.
      ; STREET: 9410 Key West Avenue
      ; CITY: Rockville
      ; STATE: Maryland
      ; COUNTRY: USA
      ; ZIP: 20850
      ; COMPUTER READABLE FORM:
      ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
      ; COMPUTER: HP Vectra 486/33
      ; OPERATING SYSTEM: MSDOS version 6.2
      ; SOFTWARE: ASCII Text
      ; CURRENT APPLICATION DATA:
      ; APPLICATION NUMBER: US/09/071,035
      ; FILING DATE:
      ; CLASSIFICATION:
      ; PRIOR APPLICATION DATA:
      ; APPLICATION NUMBER:
      ; FILING DATE:
      ; ATTORNEY/AGENT INFORMATION:
      ; NAME: A. Anders Brookes
      ; REGISTRATION NUMBER: 36,373
      ; REFERENCE/DOCKET NUMBER: PB369P2
      ; TELEPHONE: (301) 309-8504
      ; TELEFAX: (301) 309-8512
      ; INFORMATION FOR SEQ ID NO: 476:
      ; SEQUENCE CHARACTERISTICS:
      ; LENGTH: 227 amino acids
      ; TYPE: amino acid
      ; STRANDEDNESS: single
      ; TOPOLOGY: linear
      ; MOLECULE TYPE: protein
      ; US-09-071-035-476

      Query Match      46.1%; Score 41; DB 4; Length 227;
      Best Local Similarity 66.7%; Pred. No. 50;
      Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

      QY      3 DREDDFDN 11
      Db      46 NREDSFDWH 54

      RESULT 11
      US-09-543-681A-8241
      ; Sequence 8241, Application US/09543681A
      ; Patent No. 6605709
      ; GENERAL INFORMATION:
      ; APPLICANT: GARY BRETON
      ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
      ; FILE REFERENCE: 2709.1002-001
      ; CURRENT APPLICATION NUMBER: US/09/543,681A
      ; CURRENT FILING DATE: 2000-04-05
      ; PRIOR APPLICATION NUMBER: US 60/128,706
      ; PRIOR FILING DATE: 1999-04-09
      ; NUMBER OF SEQ ID NOS: 8344
      ; SEQ ID NO 8241
      ; LENGTH: 290
```

```
;
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,035
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: A. Anders Brookes
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB369P2
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 248:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 894 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
; US-09-071-035-248
;
; Query Match 46.1%; Score 41; DB 4; Length 894;
; Best Local Similarity 66.7%; Pred. No. 2.1e+02;
; Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
;
; Qy 3 DREDDFDWN 11
; Db 413 NREDSFDWH 421
;
; RESULT 14
; US-09-071-035-246
; Sequence 246, Application US/09071035
; Patent No. 6448043
; GENERAL INFORMATION:
; APPLICANT: Gil H. Choi
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 496
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,035
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: A. Anders Brookes
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB369P2
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 250:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 962 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
; US-09-071-035-250
;
; Query Match 46.1%; Score 41; DB 4; Length 962;
; Best Local Similarity 66.7%; Pred. No. 2.3e+02;
; Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
;
; Qy 3 DREDDFDWN 11
; Db 445 NREDSFDWH 453
;
; RESULT 15
; US-09-071-035-250
; Sequence 250, Application US/09071035
; Patent No. 6448043
; GENERAL INFORMATION:
; APPLICANT: Gil H. Choi
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 496
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,035
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: A. Anders Brookes
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB369P2
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 250:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 962 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
; US-09-071-035-250
;
; Query Match 46.1%; Score 41; DB 4; Length 962;
; Best Local Similarity 66.7%; Pred. No. 2.3e+02;
; Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
;
; Qy 3 DREDDFDWN 11
; Db 445 NREDSFDWH 453
```

```
;
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB369P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 246:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 962 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
; US-09-071-035-246
;
; Query Match 46.1%; Score 41; DB 4; Length 962;
; Best Local Similarity 66.7%; Pred. No. 2.3e+02;
; Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
;
; Qy 3 DREDDFDWN 11
; Db 445 NREDSFDWH 453
;
; RESULT 15
; US-09-071-035-250
; Sequence 250, Application US/09071035
; Patent No. 6448043
; GENERAL INFORMATION:
; APPLICANT: Gil H. Choi
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 496
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,035
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: A. Anders Brookes
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB369P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 250:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 962 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
; US-09-071-035-250
;
; Query Match 46.1%; Score 41; DB 4; Length 962;
; Best Local Similarity 66.7%; Pred. No. 2.3e+02;
; Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
;
; Qy 3 DREDDFDWN 11
; Db 445 NREDSFDWH 453
```

us-09-981-649a-24_copy_412_426.ra1

Fri May 21 12:56:10 2004

Search completed: May 21, 2004, 12:34:41
Job time : 0.871479 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 21, 2004, 12:31:58 ; Search time 2.05986 Seconds

(without alignments)
2031.122 Million cell updates/sec

Title: US-09-981-649A-24_COPY_412_426

Perfect score: 89

Sequence: 1 KQDREDDFWNPADR 15

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1149313 seqs, 278921704 residues

Total number of hits satisfying chosen parameters: 1149313

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/ECT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	89	100.0	502	9	US-09-981-649A-18
2	89	100.0	502	12	US-10-399-123-18
3	89	100.0	502	14	US-10-124-986-18
4	89	100.0	502	14	US-10-136-227A-18
5	89	100.0	502	14	US-10-112-881-18
6	89	100.0	537	9	US-09-981-649A-4
7	89	100.0	537	12	US-10-399-123-4
8	89	100.0	537	14	US-10-124-986-4
9	89	100.0	537	14	US-10-136-227A-4
10	89	100.0	537	14	US-10-112-881-4
11	89	100.0	551	10	US-09-866-050A-506
12	89	100.0	553	9	US-09-981-649A-6
13	89	100.0	553	9	US-09-981-649A-24
14	89	100.0	553	10	US-09-796-753-14
15	89	100.0	553	12	US-10-058-270A-102

16	89	100.0	553	12	US-10-399-123-6	Sequence 6, Appl
17	89	100.0	553	12	US-10-399-123-24	Sequence 24, Appl
18	89	100.0	553	14	US-10-124-986-6	Sequence 6, Appl
19	89	100.0	553	14	US-10-124-986-24	Sequence 24, Appl
20	89	100.0	553	14	US-10-136-227A-6	Sequence 6, Appl
21	89	100.0	553	14	US-10-136-227A-24	Sequence 24, Appl
22	89	100.0	553	14	US-10-112-881-6	Sequence 6, Appl
23	89	100.0	553	14	US-10-112-881-24	Sequence 24, Appl
24	89	100.0	553	15	US-10-295-027-494	Sequence 494, App
25	89	100.0	553	15	US-10-295-027-812	Sequence 812, App
26	89	100.0	553	15	US-10-295-027-841	Sequence 841, App
27	89	100.0	553	15	US-10-173-999-46	Sequence 46, Appl
28	89	100.0	553	16	US-10-188-832-189	Sequence 189, App
29	89	100.0	554	9	US-09-981-649A-30	Sequence 30, Appl
30	89	100.0	554	9	US-09-981-649A-32	Sequence 32, Appl
31	89	100.0	554	12	US-10-399-123-30	Sequence 30, Appl
32	89	100.0	554	12	US-10-399-123-32	Sequence 32, Appl
33	89	100.0	554	14	US-10-124-986-30	Sequence 30, Appl
34	89	100.0	554	14	US-10-124-986-32	Sequence 32, Appl
35	89	100.0	554	14	US-10-136-227A-30	Sequence 30, Appl
36	89	100.0	554	14	US-10-136-227A-32	Sequence 32, Appl
37	89	100.0	554	14	US-10-112-881-30	Sequence 30, Appl
38	89	100.0	554	14	US-10-112-881-32	Sequence 32, Appl
39	89	100.0	559	9	US-09-981-649A-28	Sequence 28, Appl
40	89	100.0	559	12	US-10-399-123-28	Sequence 28, Appl
41	89	100.0	559	14	US-10-124-986-28	Sequence 28, Appl
42	89	100.0	559	14	US-10-136-227A-28	Sequence 28, Appl
43	89	100.0	559	14	US-10-112-881-28	Sequence 28, Appl
44	84	94.4	550	15	US-10-023-634-73	Sequence 73, Appl
45	73.5	82.6	15	14	US-10-124-986-33	Sequence 33, Appl

ALIGNMENTS

RESULT 1
US-09-981-649A-18
; Sequence 18, Application US/09981649A
; Patent No. US20020132250A1
; GENERAL INFORMATION:
; APPLICANT: Ford et al.
; TITLE OF INVENTION: EGF MOTIF PROTEIN, EGFL6, MATERIALS AND METHODS
; FILE REFERENCE: 28110/37665
; CURRENT APPLICATION NUMBER: US/09/981.649A
; CURRENT FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: US 09/687,860
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 09/620,312
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: US 09/363,316
; PRIOR FILING DATE: 1999-07-28
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 18
; LENGTH: 502
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (501)
; OTHER INFORMATION: Xaa = Any Amino Acid
; NAME/KEY: misc feature
; LOCATION: (502)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-981-649A-18

Query Match 100.0%; Score 89; DB 9; Length 502;
Best Local Similarity 100.0%; Pred. No. 3.4e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 KQDREDDFWNPADR 15
Db 361 KQDREDDFWNPADR 375

US-10-399-123-18
; Sequence 18, Application US/10399123
; Publication No. US20040059098A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq et al.
; TITLE OF INVENTION: EGF MOTIF PROTEIN, EGFL6, MATERIALS AND METHODS
; FILE REFERENCE: 28110/37665
; CURRENT APPLICATION NUMBER: US/10/399,123
; PRIOR FILING DATE: 2003-04-14
; PRIOR APPLICATION NUMBER: US 09/687,860
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 09/620,312
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: US 09/363,316
; PRIOR FILING DATE: 1999-07-28
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 18
; LENGTH: 502
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(502)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-10-399-123-18

Query Match 100.0%; Score 89; DB 12; Length 502;
Best Local Similarity 100.0%; Pred. No. 3.4e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KQREDDDFDWNPADR 15
Db 361 KQREDDDFDWNPADR 375

RESULT 3
US-10-124-986-18
; Sequence 18, Application US/10124986
; Publication No. US20030036508A1
; GENERAL INFORMATION:
; APPLICANT: Ford et al.
; TITLE OF INVENTION: EGF MOTIF PROTEIN, EGFL6, MATERIALS AND METHODS
; FILE REFERENCE: 28110/37958
; CURRENT APPLICATION NUMBER: US/10/124,986
; CURRENT FILING DATE: 2002-04-17
; PRIOR APPLICATION NUMBER: US 09/981,649
; PRIOR FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: US 09/687,860
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 09/620,312
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: US 09/363,316
; PRIOR FILING DATE: 1999-07-28
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 18
; LENGTH: 502
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (501)
; OTHER INFORMATION: Xaa = Any Amino Acid
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (502)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-10-124-986-18

Query Match 100.0%; Score 89; DB 14; Length 502;
Best Local Similarity 100.0%; Pred. No. 3.4e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KQREDDDFDWNPADR 15
Db 361 KQREDDDFDWNPADR 375

RESULT 4
US-10-136-227A-18
; Sequence 18, Application US/10136227A
; Publication No. US20030165886A1
; GENERAL INFORMATION:
; APPLICANT: Ford et al.
; TITLE OF INVENTION: EGF MOTIF PROTEIN, EGFL6, MATERIALS AND METHODS
; FILE REFERENCE: 28110/38407
; CURRENT APPLICATION NUMBER: US/10/136,227A
; CURRENT FILING DATE: 2002-05-01
; PRIOR APPLICATION NUMBER: US 09/981,649
; PRIOR FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: US 09/687,860
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 09/620,312
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: US 09/363,316
; PRIOR FILING DATE: 1999-07-28
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 18
; LENGTH: 502
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (501)
; OTHER INFORMATION: Xaa = Any Amino Acid
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (502)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-10-136-227A-18

Query Match 100.0%; Score 89; DB 14; Length 502;
Best Local Similarity 100.0%; Pred. No. 3.4e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KQREDDDFDWNPADR 15
Db 361 KQREDDDFDWNPADR 375

RESULT 5
US-10-112-881-18
; Sequence 18, Application US/10112881
; Publication No. US20030166909A1
; GENERAL INFORMATION:
; APPLICANT: Ford et al.
; TITLE OF INVENTION: EGF MOTIF PROTEIN, EGFL6, MATERIALS AND METHODS
; FILE REFERENCE: 28110/38363
; CURRENT APPLICATION NUMBER: US/10/112,881
; CURRENT FILING DATE: 2002-03-29
; PRIOR APPLICATION NUMBER: US 09/981,649
; PRIOR FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: US 09/687,860
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 09/620,312
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: US 09/363,316
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: US 09/249,697
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 08/968,800

```

; PRIOR FILING DATE: 1997-11-22
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 18
; LENGTH: 502
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (501)
; OTHER INFORMATION: Xaa = Any Amino Acid
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (502)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-10-112-881-18

Query Match      100.0%; Score 89; DB 14; Length 502;
Best Local Similarity 100.0%; Pred. No. 3.4e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 KQREDDFDWNPADR 15
      |||||
Db      361 KQREDDFDWNPADR 375

RESULT 6
US-09-981-649A-4
; Sequence 4, Application US/09981649A
; Patent No. US20020132250A1
; GENERAL INFORMATION:
; APPLICANT: Ford et al.
; TITLE OF INVENTION: EGF MOTIF PROTEIN, EGFL6, MATERIALS AND METHODS
; FILE REFERENCE: 28110/37665
; CURRENT APPLICATION NUMBER: US/09/981,649A
; CURRENT FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: US 09/687,860
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 09/620,312
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: US 09/363,316
; PRIOR FILING DATE: 1999-07-28
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 537
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (503)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-981-649A-4

Query Match      100.0%; Score 89; DB 9; Length 537;
Best Local Similarity 100.0%; Pred. No. 3.6e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 KQREDDFDWNPADR 15
      |||||
Db      361 KQREDDFDWNPADR 375

RESULT 7
US-10-399-123-4
; Sequence 4, Application US/10399123
; Publication No. US20040059098A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq et al.
; TITLE OF INVENTION: EGF MOTIF PROTEIN, EGFL6, MATERIALS AND METHODS
; FILE REFERENCE: 28110/37665
; CURRENT APPLICATION NUMBER: US/10/399,123
; CURRENT FILING DATE: 2003-04-14

```

```

; PRIOR APPLICATION NUMBER: US 09/687,860
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 09/620,312
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: US 09/363,316
; PRIOR FILING DATE: 1999-07-28
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 537
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(537)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-10-399-123-4

Query Match      100.0%; Score 89; DB 12; Length 537;
Best Local Similarity 100.0%; Pred. No. 3.6e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 KQREDDFDWNPADR 15
      |||||
Db      361 KQREDDFDWNPADR 375

RESULT 8
US-10-124-986-4
; Sequence 4, Application US/10124986
; Publication No. US20030036508A1
; GENERAL INFORMATION:
; APPLICANT: Ford et al.
; TITLE OF INVENTION: EGF MOTIF PROTEIN, EGFL6, MATERIALS AND METHODS
; FILE REFERENCE: 28110/37958
; CURRENT APPLICATION NUMBER: US/10/124,986
; CURRENT FILING DATE: 2002-04-17
; PRIOR APPLICATION NUMBER: US 09/981,649
; PRIOR FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: US 09/687,860
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 09/620,312
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: US 09/363,316
; PRIOR FILING DATE: 1999-07-28
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 537
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (503)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-10-124-986-4

Query Match      100.0%; Score 89; DB 14; Length 537;
Best Local Similarity 100.0%; Pred. No. 3.6e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 KQREDDFDWNPADR 15
      |||||
Db      361 KQREDDFDWNPADR 375

RESULT 9
US-10-136-227A-4
; Sequence 4, Application US/10136227A
; Publication No. US20030165886A1
; GENERAL INFORMATION:
; APPLICANT: Ford et al.
; TITLE OF INVENTION: EGF MOTIF PROTEIN, EGFL6, MATERIALS AND METHODS

```

Fri May 21 12:56:10 2004

```

; FILE REFERENCE: 28110/38407
; CURRENT APPLICATION NUMBER: US/10/136,227A
; PRIOR FILING DATE: 2002-05-01
; PRIOR APPLICATION NUMBER: US 09/981,649
; PRIOR FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: US 09/687,860
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 09/620,312
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: US 09/363,316
; PRIOR FILING DATE: 1999-07-28
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 537
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (503)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-10-136-227A-4

Query Match 100.0%; Score 89; DB 14; Length 537;
Best Local Similarity 100.0%; Pred. No. 3.6e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KQDREDDFDWNPADR 15
Db 361 KQDREDDFDWNPADR 375

RESULT 10
US-10-112-881-4
; Sequence 4, Application US/10112881
; Publication No. US20030166909A1
; GENERAL INFORMATION:
; APPLICANT: Ford et al.
; TITLE OF INVENTION: EGF MOTIF PROTEIN, EGFL6, MATERIALS AND METHODS
; FILE REFERENCE: 28110/38363
; CURRENT APPLICATION NUMBER: US/10/112,881
; CURRENT FILING DATE: 2002-03-29
; PRIOR APPLICATION NUMBER: US 09/981,649
; PRIOR FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: US 09/687,860
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 09/620,312
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: US 09/363,316
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: US 09/249,697
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 08/968,800
; PRIOR FILING DATE: 1997-11-22
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 537
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (503)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-10-112-881-4

Query Match 100.0%; Score 89; DB 14; Length 537;
Best Local Similarity 100.0%; Pred. No. 3.6e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KQDREDDFDWNPADR 15
Db 361 KQDREDDFDWNPADR 375

RESULT 11
US-09-866-050A-506
; Sequence 506, Application US/09866050A
; Publication No. US20030040471A1
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James G.
; APPLICANT: Kumble, Krishanand D. Isolated From Skin Cells
; TITLE OF INVENTION: Compositions and Methods for Their Use
; FILE REFERENCE: 11000.1011c4U
; CURRENT APPLICATION NUMBER: US/09/866,050A
; CURRENT FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 725
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 506
; LENGTH: 551
; TYPE: PRT
; ORGANISM: Rat
US-09-866-050A-506

Query Match 100.0%; Score 89; DB 10; Length 551;
Best Local Similarity 100.0%; Pred. No. 3.7e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KQDREDDFDWNPADR 15
Db 410 KQDREDDFDWNPADR 424

RESULT 12
US-09-981-649A-6
; Sequence 6, Application US/09981649A
; Patent No. US20020132250A1
; GENERAL INFORMATION:
; APPLICANT: Ford et al.
; TITLE OF INVENTION: EGF MOTIF PROTEIN, EGFL6, MATERIALS AND METHODS
; FILE REFERENCE: 28110/37685
; CURRENT APPLICATION NUMBER: US/09/981,649A
; CURRENT FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: US 09/687,860
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 09/620,312
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: US 09/363,316
; PRIOR FILING DATE: 1999-07-28
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (357)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-981-649A-6

Query Match 100.0%; Score 89; DB 9; Length 553;
Best Local Similarity 100.0%; Pred. No. 3.7e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KQDREDDFDWNPADR 15
Db 412 KQDREDDFDWNPADR 426

RESULT 13

```

```
US-09-981-649A-24
; Sequence 24, Application US/09981649A
; Patent No. US20020132250A1
; GENERAL INFORMATION:
; APPLICANT: Ford et al.
; TITLE OF INVENTION: EGF MOTIF PROTEIN, EGFL6, MATERIALS AND METHODS
; FILE REFERENCE: 28110/37665
; CURRENT APPLICATION NUMBER: US/09/981,649A
; PRIOR FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: US 09/687,860
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 09/620,312
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: US 09/363,316
; PRIOR FILING DATE: 1999-07-28
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 24
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(553)
US-09-981-649A-24

Query Match      100.0%; Score 89; DB 9; Length 553;
Best Local Similarity 100.0%; Pred. No. 3.7e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 KQREDDFDWNPADR 15
Db      412 KQREDDFDWNPADR 426

RESULT 14
US-09-796-753-14
; Sequence 14, Application US/09796753
; Publication No. US20030027998A1
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Sean A.
; TITLE OF INVENTION: SECRETED PROTEINS AND USES THEREOF
; FILE REFERENCE: 7853-227-999
; CURRENT APPLICATION NUMBER: US/09/796,753
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 09/183,175
; PRIOR FILING DATE: 1998-10-30
; PRIOR APPLICATION NUMBER: 09/223,094
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/223,546
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/224,246
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/259,388
; PRIOR FILING DATE: 1999-02-26
; PRIOR APPLICATION NUMBER: 60/122,458
; PRIOR FILING DATE: 1999-03-01
; PRIOR APPLICATION NUMBER: 09/312,359
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 09/336,536
; PRIOR FILING DATE: 1999-06-18
; PRIOR APPLICATION NUMBER: 09/342,687
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: 09/345,464
; PRIOR FILING DATE: 1999-06-30
; PRIOR APPLICATION NUMBER: 09/365,164
; PRIOR FILING DATE: 1999-07-30
; PRIOR APPLICATION NUMBER: 09/399,723
; PRIOR FILING DATE: 1999-08-20
; PRIOR APPLICATION NUMBER: 09/409,634
; PRIOR FILING DATE: 1999-09-30
; PRIOR APPLICATION NUMBER: 09/471,179
; PRIOR FILING DATE: 1999-12-23
```

```
US-09-981-649A-24
; PRIOR APPLICATION NUMBER: 09/474,071
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 09/474,072
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 09/514,010
; PRIOR FILING DATE: 2000-02-25
; PRIOR APPLICATION NUMBER: 09/516,745
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/572,002
; PRIOR FILING DATE: 2000-05-14
; PRIOR APPLICATION NUMBER: 09/597,993
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: 09/599,596
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 09/630,334
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: 09/606,565
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: 09/606,317
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: 09/665,666
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: 09/677,751
; PRIOR FILING DATE: 2000-09-30
; NUMBER OF SEQ ID NOS: 162
; SEQ ID NO 14
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-796-753-14

Query Match      100.0%; Score 89; DB 10; Length 553;
Best Local Similarity 100.0%; Pred. No. 3.7e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 KQREDDFDWNPADR 15
Db      412 KQREDDFDWNPADR 426

RESULT 15
US-10-058-270A-102
; Sequence 102, Application US/10058270A
; Publication No. US20040029114A1
; GENERAL INFORMATION:
; APPLICANT: Mack, David H.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Afar, Daniel
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Breast Cancer, Compositions and
; FILE REFERENCE: 018501-005210US
; CURRENT APPLICATION NUMBER: US/10/058,270A
; CURRENT FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: US 60/263,965
; PRIOR FILING DATE: 2001-01-24
; PRIOR APPLICATION NUMBER: US 60/265,928
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: US 09/829,472
; PRIOR FILING DATE: 2001-04-09
; PRIOR APPLICATION NUMBER: US 60/282,698
; PRIOR FILING DATE: 2001-04-09
; PRIOR APPLICATION NUMBER: US 60/288,590
; PRIOR FILING DATE: 2001-05-04
; PRIOR APPLICATION NUMBER: US 60/294,443
; PRIOR FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 141
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 102
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-058-270A-102
```

Fri May 21 12:56:10 2004

Query Match 100.0%; Score 89; DB 12; Length 553;
Best Local Similarity 100.0%; Pred. No. 3.7e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KQDREDDFDWNPADR 15
| | | | | | | | | | | | | | | |
Db 412 KQDREDDFDWNPADR 426

Search completed: May 21, 2004, 12:44:52
Job time : 2.05986 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 21, 2004, 12:28:32 ; Search time 0.739437 Seconds

(without alignments)
1951.314 Million cell updates/sec

Title: US-09-981-649A-24_COPY_412_426

Perfect score: 89

Sequence: 1 KQREDDFDWNPADR 15

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78:*
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	89	100.0	558	2 T17324	hypothetical prote
2	48	53.9	521	2 T37504	hypothetical prote
3	47	52.8	498	2 D90371	hypothetical prote
4	46.5	52.2	875	2 I53350	karyopherin beta -
5	46.5	52.2	876	2 I52907	importin beta chai
6	46.5	52.2	876	2 S66288	nuclear pore-targe
7	45	50.6	651	2 S24609	cytoskeletal prote
8	44	49.4	330	2 S55595	G protein-coupled
9	44	49.4	435	2 E86266	protein F3F19.23 [
10	44	49.4	460	2 D84396	signal recognition
11	44	49.4	898	2 T25167	hypothetical prote
12	44	49.4	926	2 AG1860	hypothetical prote
13	43	48.3	426	2 S53399	SFH1 protein - yea
14	43	48.3	457	2 G90354	hypothetical prote
15	42	47.2	200	2 AC2392	conserved hypothet
16	42	47.2	218	2 G98291	hypothetical prote
17	42	47.2	324	2 C84225	hypothetical prote
18	42	47.2	387	2 E71711	probable o-sialogl
19	42	47.2	492	2 S10395	catalase (EC 1.11.
20	42	47.2	492	2 S48124	catalase (EC 1.11.
21	42	47.2	492	2 S10770	catalase (EC 1.11.
22	42	47.2	733	2 JC7679	dendritic cell-der
23	42	47.2	891	2 G82543	conserved hypothet
24	42	47.2	1385	2 A88554	protein C38C10.5a
25	42	47.2	1391	2 B88554	protein C38C10.5b
26	41	46.1	136	2 T46900	hypothetical prote
27	41	46.1	141	2 A70556	probable mutator M
28	41	46.1	485	2 S48650	catalase (EC 1.11.
29	41	46.1	492	2 T16969	catalase (EC 1.11.

30	41	46.1	492	2 S62696	catalase (EC 1.11.
31	41	46.1	492	2 S07124	catalase (EC 1.11.
32	41	46.1	494	2 S62897	catalase (EC 1.11.
33	41	46.1	1436	2 T14895	DNA helicase 1 - A
34	41	46.1	2747	2 B49132	fat facets (faf) s
35	40.5	45.5	479	2 A25052	fibrinogen beta ch
36	40	44.9	384	2 T24596	hypothetical prote
37	40	44.9	386	2 G87697	mannose-6-phosphat
38	40	44.9	433	2 S69999	sterigmatocystin s
39	40	44.9	433	2 T19632	hypothetical prote
40	40	44.9	473	2 T45954	hypothetical prote
41	40	44.9	510	2 H96724	F20P5.8 [imported]
42	40	44.9	580	2 T43485	hypothetical prote
43	40	44.9	621	2 I38467	low density lipopr
44	40	44.9	634	2 F90609	exodeoxyribonuclea
45	39.5	44.4	155	2 S19515	hypothetical prote

ALIGNMENTS

RESULT 1

T17324
hypothetical protein DKFP564P2063.1 - human (fragment)
C:Species: Homo sapiens (man)
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T17324
R:Dueterhoeft, A.; Lauber, J.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, September 1999
A:Reference number: Z18727
A:Accession: T17324
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-558 <DUE>
A:Cross-references: EMBL:AL117610
A:Experimental source: fetal brain; clone DKFP564P2063
C:Genetics:
A>Note: DKFP564P2063.1

Query Match 100.0%; Score 89; DB 2; Length 558;
Best Local Similarity 100.0%; Pred. No. 1.9e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1 KQREDDFDWNPADR 15
DB	417 KQREDDFDWNPADR 431

RESULT 2

T37504
hypothetical protein SPAC10F6.10 - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T37504
R:Connor, R.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
submitted to the EMBL Data Library, November 1997
A:Reference number: Z1719
A:Accession: T37504
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-521 <CON>
A:Cross-references: EMBL:AL009197; PIDN:CAA15723.1; GSPDB:GN000066; SPDB:SPAC10F6.10
A:Experimental source: strain 972h-; cosmid c10F6
C:Genetics:
A:Gene: SPDB:SPAC10F6.10
A:Map position: 1
A:Introns: 43/1; 85/1

Query Match 53.9%; Score 48; DB 2; Length 521;
Best Local Similarity 80.0%; Pred. No. 6.6;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY	4 REDDFDWNPA 13
----	-----------------

DB 31 REDDIDWNOA 40

||||| |||||

RESULT 3

D90371

hypochemical protein acsa-5 [imported] - Sulfolobus solfataricus

C:Species: Sulfolobus solfataricus

C>Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 15-Jun-2001

C:Accession: D90371

R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, H.
arrett, R.A.; Ragan, M.A.; Senses, C.W.; Van der Oost, J.

submitted to GenBank, April 2001

A:Description: Sulfolobus solfataricus complete genome.

A:Reference number: A99139

A:Accession: D90371

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-498 <KUR>

A:CROSS-references: GB:AE006641; NID:gl3815325; PIDN:AAK42227.1; GSPDB:GN00155

C:Genetics:

A:Gene: acsa-5

C:Superfamily: human SA protein; acetate-CoA ligase homology

Query Match 52.8%; Score 47; DB 2; Length 498;
Best Local Similarity 72.7%; Pred. No. 9;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 4 REDDFWNPAD 14
|||:| |||||

DB 199 RENDVFWNPAD 209

RESULT 4

I59350

karyopherin beta - rat

C:Species: Rattus norvegicus (Norway rat)

C>Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 05-Nov-1999

C:Accession: I59350

R:Radu, A.; Blobel, G.; Moore, M.S.

Proc. Natl. Acad. Sci. U.S.A. 92, 1769-1773, 1995

A:Title: Identification of a protein complex that is required for nuclear protein import

A:Reference number: I59350; MUID:95183547; PMID:7878057

A:Accession: I59350

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-875 <RES>

A:CROSS-references: GB:L38644; NID:g712838; PIDN:AAK42047.1; PID:g712839

Query Match 52.2%; Score 46.5; DB 2; Length 875;
Best Local Similarity 69.2%; Pred. No. 21;
Matches 9; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

QY 1 KQD-REDDFDWNP 12
|||:| |||||

DB 331 KQDENDDDDDWNP 343

RESULT 5

I52907

importin beta chain - human

N:Alternate names: nuclear protein import factor

C:Species: Homo sapiens (man)

C>Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 05-Nov-1999

C:Accession: I52907; A56959

R:Goerlich, D.; Kostka, S.; Kraft, R.; Dingwall, C.; Laskey, R.A.; Hartmann, E.; Prehn,
Curr. Biol. 5, 383-392, 1995

A:Title: Two different subunits of importin cooperate to recognize nuclear envelope loca

A:Reference number: I52907; MUID:95353691; PMID:7627554

A:Accession: I52907

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-876 <GOE>

A:CROSS-references: GB:L38951; NID:g893287; PIDN:AAK41763.1; PID:g893288

R:Chi, N.C.; Adam, E.J.H.; Adam, S.A.

J. Cell Biol. 130, 265-274, 1995

A:Title: Sequence and characterization of cytoplasmic nuclear protein import factor p97.

A:Reference number: A56959; MUID:95340629; PMID:7615630

A:Accession: A56959

A>Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-96, 'H', '98-705, 736-862, 'R', 864-876 <CHI>

A:CROSS-references: GB:L39793; NID:g755649

A:Experimental source: HeLa cell mRNA

C:Genetics:

A:Gene: GDB:KPNB1

A:CROSS-references: GDB:6053802

Query Match 52.2%; Score 46.5; DB 2; Length 876;
Best Local Similarity 69.2%; Pred. No. 21;
Matches 9; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

QY 1 KQD-REDDFDWNP 12
|||:| |||||

DB 332 KQDENDDDDDWNP 344

RESULT 6

S66288

nuclear pore-targeting complex protein, 97K - mouse

C:Species: Mus musculus (house mouse)

C>Date: 15-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 05-Nov-1999

C:Accession: S66288; S74308

R:Imamoto, N.; Shimamoto, T.; Kose, S.; Takao, T.; Tachibana, T.; Matsubae, M.; Sekimoto

FEBS Lett. 368, 415-419, 1995

A:Title: The nuclear pore-targeting complex binds to nuclear pores after association wi

A:Reference number: S66288; MUID:95361913; PMID:7635189

A:Accession: S66288

A:Molecule type: mRNA

A:Residues: 1-876 <IMA>

A:CROSS-references: EMBL:D45836; NID:g871889; PIDN:BAAB0273.1; PID:d1008868; PID:g87189

A>Note: In the authors' translation residues 637-657 do not match the nucleotide sequen

A:Accession: S74308

A:Molecule type: protein

A:Residues: 74-92; 212-221; 377-389; 860-867 <IMM>

Query Match 52.2%; Score 46.5; DB 2; Length 876;
Best Local Similarity 69.2%; Pred. No. 21;
Matches 9; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

QY 1 KQD-REDDFDWNP 12
|||:| |||||

DB 332 KQDENDDDDDWNP 344

RESULT 7

S24609

cytoskeletal protein - Euglena gracilis

C:Species: Euglena gracilis

C>Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 01-Dec-2000

C:Accession: S24609; A43417

R:Bouck, G.B.

submitted to the EMBL Data Library, June 1992

A:Reference number: S24609

A:Accession: S24609

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-651 <BOU>

A:CROSS-references: EMBL:Z13962; NID:gl8403; PID:gl8404

R:Marris, J.A.; Bouck, G.B.

J. Cell Biol. 118, 1465-1475, 1992

A:Title: The two major membrane skeletal proteins (articuline) of Euglena gracilis defu

A:Reference number: A43417; MUID:92394973; PMID:1522117

A:Accession: A43417

A>Status: preliminary; not compared with conceptual translation

Fri May 21 12:56:11 2004

A:Molecule type: nucleic acid
A:Residues: 1-145,'T',147-187,'T',189-348,'T',350-651 <MAR>
A>Note: sequence extracted from NCBI backbone (NCBIP:113032)

Query Match 50.6%; Score 45; DB 2; Length 651;
Best Local Similarity 77.8%; Pred. No. 26; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 2; Indels 2; Gaps 0;

QY 4 REDDFDWWNP 12
| | | | |
Db 631 RNDDFDWWPP 639

RESULT 8
S55595
G protein-coupled receptor 74 - equine herpesvirus 2
C:Species: equine herpesvirus 2
C>Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 17-Mar-2000
C:Accession: S55595
R:Telford, E.A.R.; Watson, M.S.; Aird, H.C.; Perry, J.; Davison, A.J.
J. Mol. Biol. 249, 520-528, 1995
A:Title: The DNA sequence of equine herpesvirus 2.
A:Reference number: S55594; MUID:95302501; PMID:7783207
A:Accession: S55595
A>Status: preliminary; nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-330 <TEL>
A:Cross-references: GB:U20824; NID:G695172; PIDN:AAC13861.1; PID:G695246
C:Superfamily: vertebrate rhodopsin
C:Keywords: G protein-coupled receptor

Query Match 49.4%; Score 44; DB 2; Length 330;
Best Local Similarity 66.7%; Pred. No. 17; Indels 0; Gaps 0;
Matches 6; Conservative 2; Mismatches 1; Indels 1; Gaps 0;

QY 6 DDFDWWNPAD 14
| | | | |
Db 26 DDYDWNSSD 34

RESULT 9
E86266
protein F3F19.23 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: E86266
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizlar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: E86266
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-435 <STO>
A:Cross-references: GB:AE005172; NID:G4850403; PIDN:AAD31073.1; GSPDB:GN00141
C:Genetics:
A:Gene: F3F19.23
A:Map position: 1

Query Match 49.4%; Score 44; DB 2; Length 435;
Best Local Similarity 40.0%; Pred. No. 23; Indels 0; Gaps 0;
Matches 5; Conservative 4; Mismatches 5; Indels 5; Gaps 0;

QY 1 KQREDDFDWWNPAD 15
| | | | |
Db 363 RTEKRDQFDWSKKDR 377

RESULT 10
D84396
signal recognition particle [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C:Accession: D84396
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky,
Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jabl
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; L
A:Title: Genome sequence of Halobacterium species NRC-1.
A:Reference number: A84160; MUID:20504483; PMID:11016950
A:Accession: D84396
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-460 <STO>
A:Cross-references: GB:AE004437; NID:G10581860; PIDN:AG20536.1; GSPDB:GN00138
C:Genetics:
A:Gene: srp54
C:Superfamily: signal recognition particle 54K protein

Query Match 49.4%; Score 44; DB 2; Length 460;
Best Local Similarity 42.9%; Pred. No. 25; Indels 0; Gaps 0;
Matches 6; Conservative 5; Mismatches 3; Indels 3; Gaps 0;

QY 1 KQREDDFDWWNPAD 14
| | | | |
Db 306 EETQDDDDWDPEP 319

RESULT 11
T25167
hypothetical protein T23D8.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T25167
R:Wild, A.
submitted to the EMBL Data Library, October 1996
A:Reference number: Z19989
A:Accession: T25167
A>Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-898 <WIL>
A:Cross-references: EMBL:Z81128; PIDN:CA803403.1; GSPDB:GN00019; CESP:T23D8.4
A:Experimental source: clone T23D8
C:Genetics:
A:Gene: CESP:T23D8.4
A:Map position: 1
A:Introns: 34/2, 112/2; 407/3; 780/2

Query Match 49.4%; Score 44; DB 2; Length 898;
Best Local Similarity 46.2%; Pred. No. 54; Indels 0; Gaps 0;
Matches 6; Conservative 3; Mismatches 4; Indels 4; Gaps 0;

QY 3 DREDDFDWWNPAD 15
| | | | |
Db 274 EEDDDADWTVPNR 286

RESULT 12
AG1860
hypothetical protein all0432 [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AG1860
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana

Fri May 21 12:56:11 2004

```

A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AG1860
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-926 <KUR>
A;Cross-references: GB:BA000019; PIDN:BA872390.1; PID:gl7129777; GSPDB:GN00179
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: all0432

Query Match      49.4%; Score 44; DB 2; Length 926;
Best Local Similarity 77.8%; Pred. No. 56;
Matches      7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      6 DDFDWNPAD 14
Db      367 DDADWQPAD 375

RESULT 13
S53399
SFH1 protein - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein L8543.4; protein YLR321c
C:Species: Saccharomyces cerevisiae
C;Date: 05-May-1995 #sequence_revision 01-Sep-1995 #text change 05-Nov-1999
C;Accession: S53399
A;Molecule type: DNA
A;Residues: 1-426 <DUZ>
A;Cross-references: EMBL:U20618; NID:g2258165; PID:g662128; GSPDB:GN00012; MIPS:YLR321c
A;Experimental source: strain S288C (AB972)
C;Genetics:
A;Gene: SGD:SFH1; MIPS:YLR321c
A;Cross-references: MIPS:YLR321c; SGD:S0004313
A;Map position: 12R

Query Match      48.3%; Score 43; DB 2; Length 426;
Best Local Similarity 56.2%; Pred. No. 33;
Matches      9; Conservative 2; Mismatches 3; Indels 2; Gaps 1;

QY      2 QDR--EDDFWNPADR 15
Db      288 QDRFFEDNFQNLNDK 303

RESULT 14
G90354
hypothetical protein acsA-3 [imported] - Sulfolobus solfataricus
C;Species: Sulfolobus solfataricus
C;Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001
C;Accession: G90354
R;She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P.
arrett, R.A.; Ragan, M.A.; Senses, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A;Description: Sulfolobus solfataricus complete genome.
A;Reference number: A99139
A;Accession: G90354
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-457 <KUR>
A;Cross-references: GB:AE006641; NID:g13815165; PIDN:AAK42094.1; GSPDB:GN00155
C;Genetics:
A;Gene: acsA-3

Query Match      48.3%; Score 43; DB 2; Length 457;
Best Local Similarity 63.6%; Pred. No. 36;
Matches      7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```

```

QY      4 REDDFDWNPAD 14
Db      199 RDSDFVFNPAD 209

RESULT 15
AC2992
conserved hypothetical protein Atu3541 [imported] - Agrobacterium tumefaciens (strain C
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C;Accession: AC2992
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo,
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McCllel
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; MUID:21608550; PMID:11743193
A;Accession: AC2992
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-200 <KUR>
A;Cross-references: GB:AE008689; PIDN:AA44353.1; PID:gl7741946; GSPDB:GN00187
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: Atu3541
A;Map position: linear chromosome

Query Match      47.2%; Score 42; DB 2; Length 200;
Best Local Similarity 61.5%; Pred. No. 20;
Matches      8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY      1 QDRDDDFDWNPA 13
Db      104 EQGDDDKWLNPA 116

Search completed: May 21, 2004, 12:33:58
Job time : 2.73944 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 21, 2004, 12:27:01 ; Search time 0.528169 Seconds
(without alignments)
1478.792 Million cell updates/sec

Title: US-09-981-649A-24_COPY_412_426

Perfect score: 89
Sequence: 1 KQDREDDFDWNPADR 15

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	50	56.2	243	1 RWD1_HUMAN	Q9h446 homo sapien
2	46.5	52.2	875	1 IMB1_RAT	P52296 rattus norv
3	46.5	52.2	876	1 IMB1_HUMAN	Q14974 homo sapien
4	46.5	52.2	876	1 IMB1_MOUSE	P70168 mus musculu
5	45.5	51.1	1914	1 RLF_HUMAN	Q13129 homo sapien
6	45	50.6	859	1 RLF_HUMAN	Q8nf97 homo sapien
7	44	49.4	460	1 SR54_HUMAN	Q9hmn5 halobacteri
8	44	49.4	898	1 IF38_CAEEL	Q02328 caenorhabdi
9	44	49.4	4074	1 PKHD_HUMAN	Q8tcz9 homo sapien
10	43	48.3	135	1 LECG_BOTUR	P83519 bothrops ja
11	43	48.3	147	1 MGN_DROME	P49028 drosophila
12	43	48.3	723	1 FCP1_SCHPO	Q9p376 schizosacch
13	43	48.3	881	1 PHFE_MOUSE	Q8d4b9 mus musculu
14	43	48.3	888	1 PHFE_HUMAN	Q94880 homo sapien
15	42	47.2	387	1 GCP_RICPR	Q9zeaz rickettsia
16	42	47.2	492	1 CAT1_GOSHI	P17598 gossypium h
17	42	47.2	492	1 CAT1_MAIZE	P18122 zea mays (m
18	42	47.2	492	1 CAT1_AVIME	Q9axh0 avicennia m
19	42	47.2	492	1 CAT1_SOLAP	Q24339 soldanella
20	42	47.2	741	1 Z288_HUMAN	Q9hc78 homo sapien
21	42	47.2	1342	1 Z335_HUMAN	Q9h422 homo sapien
22	42	47.2	1391	1 YLD5_CAEEL	Q03570 caenorhabdi
23	41	46.1	158	1 LEGC_TRIST	Q9y9p1 trimeresuru
24	41	46.1	485	1 CAT1_NICPL	P49317 nicotiana p
25	41	46.1	492	1 CAT1_HORVU	P55307 hordeum vul
26	41	46.1	492	1 CAT2_WHEAT	P55313 triticum ae
27	41	46.1	492	1 CAT3_NICPL	P49315 nicotiana p
28	41	46.1	492	1 CAT1_POBA	P07145 ipomoea bat
29	41	46.1	492	1 CAT2_ORYSA	P55309 oryza sativ
30	41	46.1	494	1 CAT2_HORVU	P55308 hordeum vul
31	41	46.1	509	1 HOV1_YARLI	Q99160 yarrowia li
32	41	46.1	622	1 YAA1_HUMAN	Q9upw0 homo sapien
33	41	46.1	719	1 IF39_TOBAC	P56821 nicotiana t

34	41	46.1	749	1 SPOT_SPICI	O34098 spiroplasma
35	41	46.1	898	1 Z03_CANFA	O62683 canis famil
36	41	46.1	933	1 Z03_HUMAN	O95049 homo sapien
37	41	46.1	1436	1 WRN_XENLA	O95330 xenopus lae
38	41	46.1	1562	1 YFRD_SCHPO	Q9ut43 schizosacch
39	41	46.1	2778	1 FAF_DROME	P55824 drosophila
40	40.5	45.5	477	1 FIBB_PETMA	P02678 petromyzon
41	40.5	45.5	527	1 RRP_TPMV	Q9ws39 tupaia para
42	40.5	45.5	603	1 YTH3_CAEEL	P54002 caenorhabdi
43	40	44.9	243	1 RWD1_MOUSE	Q9cdk7 mus musculu
44	40	44.9	243	1 RWD1_RAT	Q99nd9 rattus norv
45	40	44.9	384	1 CGB3_CAEEL	Q10654 caenorhabdi

ALIGNMENTS

RESULT 1
RWD1_HUMAN STANDARD; PRT; 243 AA.
AC Q9H416; Q9Y313; Q9Y6B3;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE RWD domain containing protein 1 (CGI-24) (PTD013).
GN RWDD1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
RX MEDLINE=2072150; PubMed=10810093;
RA Lai C.-H., Chou C.-Y., Ch'ang L.-Y., Liu C.-S., Lin W.-C.;
RT "Identification of novel human genes evolutionarily conserved in
CAenorhabditis elegans by comparative proteomics.";
RL Genome Res. 10:703-713(2000).
[2]
SEQUENCE FROM N.A.
RC TISSUE=Pituitary tumor;
RA Song H., Peng Y., Dai M., Huang Q., Mao Y., Zhang Q., Mao M., Fu G.,
LAO M., Chen J., Hu R.;
[3]
Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RA Babbage A.;
[4]
Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
SEQUENCE FROM N.A.
RC TISSUE=Uterus;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,
Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
Brownstein M.J., Udell I.B., Toshyuki S., Carninci P., Prange C.,
Raha S.S., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
Burtfield V.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16999-16903 (2002).
CC -!- SIMILARITY: Contains 1 RWD domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; AF132958; AAD27733.1; -;
DR EMBL; AF092134; AAD40376.1; -;
DR EMBL; AL121953; CAC14427.1; -;
DR EMBL; BC015802; AAH15802.1; -;
DR Genew; HGNC:20993; RWD1.
DR InterPro; IPR006575; RWD.
DR Pfam; PF05773; RWD; 1.
DR SMART; SM00591; RWD; 1.
DR PROSITE; PS0908; RWD; 1.
FT DOMAIN 10 114
FT REPEAT 120 130
FT CONFLICT 139 139
FT SEQUENCE 243 AA; 27939 MW; A7AABD457136BADA CRC64;
SQ

Query Match 56.2%; Score 50; DB 1; Length 243;
Best Local Similarity 69.2%; Pred. No. 0.79;
Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 QDREDDFDWNPAD 14

Db 224 EDDDDPDYNPAD 236

RESULT 2

IMB1_RAT STANDARD; PRT; 875 AA.
AC P52236;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Importin beta-1 subunit (Karyopherin beta-1 subunit) (Nuclear factor
DE P97).
GN KENB1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=Buffalo; TISSUE=Liver;
RX MEDLINE=95183547; PubMed=7878057;
RA Radu A., Blobel G., Moore M.M.;
RT "Identification of a protein complex that is required for nuclear
RT protein import and mediates docking of import substrate to distinct
RT nucleoporins.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:1769-1773 (1995).
RN [2]
RP SUBCELLULAR LOCATION, AND BINDING TO NUCLEOPORINS.
RX MEDLINE=95327681; PubMed=7604027;
RA Morioanu J., Hijikata M., Blobel G., Radu A.;
RT "Mammalian karyopherin alpha 1 beta and alpha 2 beta heterodimers:
RT alpha 1 or alpha 2 subunit binds nuclear localization signal and
RT beta subunit interacts with peptide repeat-containing nucleoporins.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:6532-6536 (1995).
CC -!- FUNCTION: Required for nuclear protein import and mediates docking
CC of import substrate to distinct nucleoporins. Serves as a receptor
CC for nuclear localization signals.
CC -!- SUBUNIT: FORMS A COMPLEX WITH AN IMPORTIN ALPHA SUBUNIT.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic and nuclear envelope.
CC -!- SIMILARITY: Belongs to the importin beta family.
CC -!- SIMILARITY: Contains 1 importin N-terminal domain.
CC -!- SIMILARITY: Contains 8 HEAT repeats.
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; L38644; AAC42047.1; -;
DR PIR; I59350; I59350.
DR InterPro; IPR008938; ARM.
DR InterPro; IPR000225; Armadillo.
DR InterPro; IPR000357; HEAT.
DR InterPro; IPR001494; Importinb N.
DR Pfam; PF00514; Armadillo_beg; 1.
DR Pfam; PF03810; IBN_NT; 1.
DR PROSITE; PS50077; HEAT REPEAT; 1.
DR PROSITE; PS50166; IMPORTIN B_NT; 1.
KW Transport; Protein transport; Repeat; Nuclear protein.
FT DOMAIN 21 100 IMPORTIN N-TERMINAL.
FT REPEAT 123 162 HEAT 1.
FT REPEAT 167 206 HEAT 2.
FT REPEAT 212 249 HEAT 3.
FT REPEAT 316 356 HEAT 4.
FT REPEAT 361 398 HEAT 5.
FT REPEAT 403 440 HEAT 6.
FT DOMAIN 328 341 IAB DOMAIN.
FT DOMAIN 333 418 RAN-GTP BINDING (BY SIMILARITY).
FT REPEAT 601 640 HEAT 7.
FT REPEAT 686 725 HEAT 8.
SQ SEQUENCE 875 AA; 97123 MW; 9CD77A05744014C4 CRC64;

Query Match 52.2%; Score 46.5; DB 1; Length 875;

Best Local Similarity 69.2%; Pred. No. 11;

Matches 9; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

QY 1 KQDREDDFDWNP 12

Db 331 KQDREDDFDWNP 343

RESULT 3

IMB1_HUMAN STANDARD; PRT; 876 AA.
ID IMB1_HUMAN
AC Q14974; Q14637; Q96J27;
DT 01-NOV-1997 (Rel. 35, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Importin beta-1 subunit (Karyopherin beta-1 subunit) (Nuclear factor
DE P97) (Importin 90).
GN KENB1 OF NTF97.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC TISSUE=Brain;
RX MEDLINE=95340629; PubMed=7615630;
RA Chi N.C., Adam E.J.H., Adam S.A.;
RT "Sequence and characterization of cytoplasmic nuclear protein import
RT factor p97.";
RL J. Cell Biol. 130:265-274 (1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=95353691; PubMed=7627554;
RA Gerlich D., Kostka S., Kraft R., Dingwall C., Laskey R.A.,
RA Hartmann E., Prehn S.;
RT "Two different subunits of importin cooperate to recognize nuclear
RT localization signals and bind them to the nuclear envelope.";
RL Curr. Biol. 5:383-392 (1995).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain, and Eye;
RX MEDLINE=22388257; PubMed=12477932;

```

InterPro; IPR000225; Armadillo
InterPro; IPR000357; HEAT
InterPro; IPR001494; Importinb N.
Pfam; PF00514; Armadillo seg. 1.
Pfam; PF03810; IBN NT; 1.
PROSITE; PS50077; HEAT REPEAT; 1.
PROSITE; PS50166; IMPORTIN_B_NT; 1.
Transport; Protein transport; Repeat; Nuclear protein; 3D-structure.
DOMAIN 21 101 IMPORTIN N-TERMINAL.
REPEAT 124 163 HEAT 1.
REPEAT 168 207 HEAT 2.
REPEAT 213 250 HEAT 3.
REPEAT 317 357 HEAT 4.
REPEAT 362 399 HEAT 5.
REPEAT 404 441 HEAT 6.
REPEAT 429 467 IAB BINDING.
DOMAIN 329 342 IAB BINDING.
DOMAIN 334 419 RAN-GTP BINDING.
REPEAT 602 641 HEAT 7.
REPEAT 687 726 HEAT 8.
DOMAIN 337 341 POLY-ASP.
CONFLICT 97 97 Q -> H (IN REF. 1).
CONFLICT 200 200 N -> NA (IN REF. 4).
CONFLICT 863 863 T -> R (IN REF. 1).
SEQUENCE 876 AA; 97169 MW; F3B8B73E7E51639 CRC64;

Query Match 52.2%; Score 46.5; DB 1; Length 876;
Best Local Similarity 69.2%; Pred. No. 12;
Matches 9; Conservative 1; Mismatches 2; Indels 1; Gaps

QY 1 KQD-REDDFWNP 12
||| :||| |||
332 KQDENDDDDWNP 344

DB

RESULT 4
IMBI MOUSE
ID IMBI_MOUSE STANDARD; PRT; 876 AA.
AC P70168; Q62117;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Importin beta-1 subunit (Karyopherin beta-1 subunit) (Nuclear factor
DE P97) (Pore targeting complex 97 kDa subunit) (PTAC97) (SCG).
OS KPNB1 OR IMPN8.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
X1 [1]
RN SEQUENCE FROM N.A.
RC MESSAGE=Lung;
RC TISSUE=Lung;
RC TISSUE=96411672; PubMed=8812441;
RA Matsuda Y., Hamatani K., Itoh M., Takahashi E.-I., Araki R., Abe M. &
RA "Localization of the importin-beta gene to mouse chromosome 11D and
RT rat chromosome 10q32.1";
RL Genomics 36:213-215(1996).
RN [2]
RN SEQUENCE FROM N.A., AND SEQUENCE OF 74-92; 212-221; 377-389 AND
RP 860-867.
RC STRAIN=C57BL/6; TISSUE=Thymus;
RC TISSUE=95361913; PubMed=7635189;
RA Imamoto N., Shimamoto T., Kose S., Takao T., Tachibana T.,
RA Matsubae M., Sekimoto T., Shimonishi Y., Yoneda Y.
RA "The nuclear pore-targeting complex binds to nuclear pores after
RT association with a karyophilin.";
RL FEBS Lett. 368:415-419(1995).
CC -1- FUNCTION: Required for nuclear protein import and mediates dock-
CC ing of import substrate to distinct nucleoporins. Serves as a receptor
CC for nuclear localization signals.
CC -1- SUBUNIT: FORMS A COMPLEX WITH AN IMPORTIN ALPHA SUBUNIT.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic and nuclear envelope (By
CC similarity).
CC -1- SIMILARITY: Belongs to the importin beta family.

```

Fri May 21 12:56:11 2004

```
CC -1- SIMILARITY: Contains 1 importin N-terminal domain.
CC -1- SIMILARITY: Contains 8 HEAT repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D67015; BAA11034.1; -.
CC EMBL: D45836; BAA08273.1; -.
CC EIR: S66288; S66288.
CC MGD: MGI:107532; Kpnbl.
CC InterPro: IPR008938; ARM.
CC InterPro: IPR000225; Armadillo.
CC InterPro: IPR000357; HEAT.
CC InterPro: IPR001494; Importinb N.
CC Pfam: PF00514; Armadillo_seg; 1.
CC Pfam: PF03810; IEN_NT; 1.
CC PROSITE: PS50077; HEAT_REPEAT; 1.
CC PROSITE: PS50166; IMPORTIN_B_NT; 1.
CC Transport; Protein transport; Repeat; Nuclear protein.
KW DOMAIN 21 101
FT REPEAT 124 163
FT REPEAT 168 207
FT REPEAT 213 250
FT REPEAT 317 357
FT REPEAT 362 399
FT REPEAT 404 441
FT REPEAT 329 342
FT DOMAIN 334 419
FT REPEAT 602 641
FT REPEAT 687 726
FT REPEAT 337 341
FT DOMAIN 388 388
FT CONFLICT 876 AA; 97151 MW; B540764C526419E9 CRC64;
SQ SEQUENCE 876 AA; 97151 MW; B540764C526419E9 CRC64;

Query Match 52.2%; Score 46.5; DB 1; Length 876;
Best Local Similarity 69.2%; Pred. No. 12;
Matches 9; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

QY 1 KQD-REDDFDWNP 12
Db 332 KQDENDDDDWNP 344
|||||:|||||
Query Match 51.1%; Score 45.5; DB 1; Length 1914;
Best Local Similarity 50.0%; Pred. No. 39;
Matches 7; Conservative 5; Mismatches 1; Indels 1; Gaps 1;

ID_RLP_HUMAN STANDARD; PRT; 1914 AA.
AC Q13129; Q9NU60;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Zinc finger protein Rlf (Rearranged L-myc fusion gene protein) (Zn-15
DE related protein).
GN RLF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. PubMed=8545128;
RX MEDLINE=9613723;
RA Makela T.P., Hellsten E., Vesa J., Hirvonen H., Palotie A.,
RA Peltonen L., Alitalo K.;
RT "The rearranged L-myc fusion gene (RLF) encodes a Zn-15 related zinc
RT finger protein.";
RL Oncogene 11:2699-2704(1995).
RN [2]
RP SEQUENCE OF 317-1914 FROM N.A.
RA Donnelly S.;
```

```
Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
-1- FUNCTION: May function as a transcription factor.
-1- SUBCELLULAR LOCATION: Nuclear (Probable).
-1- TISSUE SPECIFICITY: Widely expressed in fetal and adult tissues.
-1- DISEASE: In some small cell lung carcinoma (SCLC) cell lines,
there is an intrachromosomal rearrangements at 1p32 fusing the
first exon of the RLF gene with L-myc.
-1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
FINGER PROTEINS.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL: U23177; AAC50396.1; -.
EMBL: AL050341; CAB81608.1; -.
DR Genew; HGNC:10025; RLF.
DR MIM; 180610; -.
DR GO; GO:0003700; F:transcription factor activity; TAS.
DR GO; GO:0008270; P:zinc ion binding; TAS.
DR InterPro: IPR007087; Znf_C2H2.
DR Pfam; PF00096; zf-C2H2; 12.
DR SMART; SM00385; Znf_C2H2; 14.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 14.
DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 11.
KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW Nuclear protein; Repeat.
FT ZN_FING 582 604
FT ZN_FING 671 696
FT ZN_FING 714 736
FT ZN_FING 742 766
FT ZN_FING 771 795
FT ZN_FING 801 825
FT ZN_FING 954 979
FT ZN_FING 1127 1152
FT ZN_FING 1172 1195
FT ZN_FING 1310 1335
FT ZN_FING 1362 1387
FT ZN_FING 1407 1432
FT ZN_FING 1444 1469
FT ZN_FING 1549 1574
FT ZN_FING 1546 1548
FT CONFLICT 1914 AA; 217899 MW; 1E7529139F6528AA CRC64;
SQ SEQUENCE 1914 AA; 217899 MW; 1E7529139F6528AA CRC64;

Query Match 51.1%; Score 45.5; DB 1; Length 1914;
Best Local Similarity 50.0%; Pred. No. 39;
Matches 7; Conservative 5; Mismatches 1; Indels 1; Gaps 1;

QY 2 QDREDDFDWNPAD 14
Db 1781 EBKEDFDWDPSE 1794
:::|||||:::
RESULT 6
MRIP_HUMAN STANDARD; PRT; 859 AA.
ID MRIP_HUMAN STANDARD; PRT; 859 AA.
AC Q9NF99; O81UF5; Q9Y3V4;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Rab effector MyRIP (Myosin-VIIa- and Rab-interacting protein)
DE (Exophilin 8) (Slp homolog lacking C2 domains-c).
GN MYRIP OR SLAC2C.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., INTERACTION WITH RAB27A AND MYO7A, AND SUBCELLULAR
```

RP LOCATION.
RC MEDLINE=21987934; PubMed=11964381;
RX El-Amraoui A., Schom J.-S., Kuesel-Andermann P., Blanchard S.,
RA Desnos C., Henry J.-P., Wolfram U., Darchen F., Petit C.;
RT "MYRIP, a novel Rab effector, enables myosin VIIa recruitment to
RT retinal melanosomes.";
RL EMBO Rep. 3:463-470(2002).
RL [2].
RP SEQUENCE FROM N.A., AND INTERACTION WITH MYO5A AND MYO7A.
RX MEDLINE=22302022; PubMed=12221080;
RA Fukuda M., Kuroda T.S.;
RT "Slac2-c (synaptotagmin-like protein homologue lacking C2 domains-c), and
RT a novel linker protein that interacts with Rab27, myosin Va/VIIa, and
RT actin.";
RL J. Biol. Chem. 277:43096-43103(2002).
RL [3].
RP SEQUENCE OF 699-859 FROM N.A.
RC TISSUE=Uterus;
RA Wambutt R., Heubner D., Mewes H.-W., Gassenhuber J., Wiemann S.;
RA Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Rab effector protein involved in melanosome transport.
CC Serves as link between melanosome-bound RAB27A and the motor
CC proteins MYO5A and MYO7A. May link RAB27A-containing vesicles to
CC actin filaments (By similarity).
CC -!- SUBUNIT: Binds MYO5A, MYO7A and F-actin (By similarity). Binds
CC RAB27A that has been activated by GTP-binding via its N-terminus.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic; in pre- and post-synaptic areas
CC in photoreceptor cells and in the basal microvilli of retinal
CC pigment epithelium cells. Associated with melanosomes. Colocalizes
CC with actin filaments (By similarity).
CC -!- TISSUE SPECIFICITY: Detected in brain, skin, heart, adrenal
CC medulla, pancreas, intestine, liver, kidney, muscle and testis.
CC -!- SIMILARITY: Contains 1 FYVE-type zinc finger.
CC -!- SIMILARITY: Contains 1 Rab-binding domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF396687; AAA43954.1; -
CC EMBL; AB083783; BAC15555.1; -
CC EMBL; AL050090; CAB43262.1; -
CC Genew; HGNC:19156; MYRIP.
CC PROSITE; PS0916; RABBD, 1.
KW Actin-binding; Metal-binding; Zinc; Repeat; Zinc-finger.
FT DOMAIN 4 124 RAB-BINDING.
FT ZN FING 63 105 FYVE-TYPE.
FT DOMAIN 143 560 MYOSIN BINDING.
FT DOMAIN 495 856 ACTIN BINDING.
FT CONFLICT 571 571 E -> D (IN REF. 2).
SQ SEQUENCE 859 AA; 95720 MW; C165B8EBE4FF8078 CRC64;
Query Match 50.6%; Score 45; DB 1; Length 859;
Best Local Similarity 63.6%; Pred. No. 19;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY 3 DREDDFDWNPDA 13
Db 396 DSEEDFDWSEA 406
||:||||:|
STANDARD; PRT; 460 AA.
RESULT 7
SR54 HALN1
ID -SR54 HALN1
AC QSHMN5; 2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Signal recognition 54 kDa protein (SRP54).
GN SRP54 OR VNG2459G.
OS Halobacterium sp. (strain NRC-1 / ATCC 700922 / JCM 11081).
OC Archaea; Euryarchaeota; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxID=64091;
RN [1].
RP SEQUENCE FROM N.A.
RX MEDLINE=20504483; PubMed=11016950;
RA NG W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Shrogha J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA Leithausen B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Ienbarger T.A., Beck R.F., Pohlischroder M., Spudich J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Riley M., Dennis P.P., Omer A.D.,
RA Ebhardt H., Lowe T.M., Liang P., Hood L., DasSarma S.;
RA "Genome sequence of Halobacterium species NRC-1.";
RA Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
RL -!- FUNCTION: Binds to the signal sequence of presecretory protein
CC when they emerge from the ribosomes (By similarity).
CC -!- SUBUNIT: Archaeal signal recognition particle consists of a 7S RNA
CC molecule of 300 nucleotides and two protein subunits: SRP54 and
CC SRP19 (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- DOMAIN: Has a two domain structure: the G-domain binds GTP; the M-
CC domain binds the 7S RNA in presence of SRP19 and also binds the
CC signal sequence (By similarity).
CC -!- SIMILARITY: Belongs to the GTP-binding SRP family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE005125; AAG20536.1; -
CC FIR; D84396; D84396.
CC HSSP; O07347; 2FFH.
CC HAVAP; MF 00306; -; 1.
CC InterPro; IPR003593; AAA ATPase.
CC InterPro; IPR000897; SRP54.
CC InterPro; IPR004125; SRP54_SBP.
CC Pfam; PF00448; SRP54; 1.
CC Pfam; PF02881; SRP54_N; 1.
CC Pfam; PF02978; SRP_SBP; 1.
CC ProDom; PD000819; SRP54; 1.
CC SMART; SM00382; AAA; 1.
CC PROSITE; PS00300; SRP54; FALSE NEG.
KW Signal recognition particle; GTP-binding; RNA-binding;
FT DOMAIN 1 289 Complete proteome.
FT DOMAIN 290 460 G-DOMAIN (BY SIMILARITY).
FT NP_BIND 104 111 M-DOMAIN (BY SIMILARITY).
FT NP_BIND 184 188 GTP (BY SIMILARITY).
FT NP_BIND 242 245 GTP (BY SIMILARITY).
FT DOMAIN 449 458 POLY-GLY.
SQ SEQUENCE 460 AA; 50149 MW; 8361B782B651352E CRC64;
Query Match 49.4%; Score 44; DB 1; Length 460;
Best Local Similarity 42.9%; Pred. No. 14;
Matches 6; Conservative 5; Mismatches 3; Indels 0; Gaps 0;
QY 1 KQDREDDFDWNPAD 14
Db 306 EETQDDDEDWDPED 319
||:||||:|
STANDARD; PRT; 898 AA.
RESULT 8
IF38 CAEEL
ID -IF38 CAEEL

```
AC Q02328;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probable eukaryotic translation initiation factor 3 subunit 8 (eIF3
DE p110) (eIF3c).
GN T23D8.4.
OS Caenorhabditis elegans.
OS Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RC Wild A.;
RA Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Binds to the 40S ribosome and promotes the binding of
CC methionyl-tRNAI and mRNA (By similarity).
CC -!- SUBUNIT: eIF-3 is composed of at least 10 different subunits (By
CC similarity).
CC -!- SIMILARITY: Contains 1 PCI domain.
CC -!- SIMILARITY: BELONGS TO THE EIF3S8 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z81128; CAB03403.1; -
CC PIR; T25167; T25167.
CC WormPep; T23D8.4; CE18958.
CC InterPro; IPR008504; EIP3C.C.
CC InterPro; IPR008905; EIP3C.N.
CC InterPro; IPR000717; PCI.
CC Pfam; PF05469; eIF3C.C; 1.
CC Pfam; PF05470; eIF3C.N; 1.
CC Pfam; PF01399; PCI; 1.
CC SMART; SM00088; PINT; 1.
CC Initiation factor; Protein biosynthesis.
KW SEQUENCE 898 AA; 103842 MW; D8D36F89CD2FDA CRC64;
SQ
QY Query Match 49.4%; Score 44; DB 1; Length 898;
Best Local Similarity 46.2%; Pred. No. 29;
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
Db 274 EEDDDADWTGVNR 286
: ||| |||
3 DREDDDFMNPADR 15
PRT; 4074 AA.
ID PKHD HUMAN STANDARD; PRT; 4074 AA.
AC Q8TCZ9;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Polycystic kidney and hepatic disease 1 precursor (Fibrocystin)
DE (Polyductin) (Tigmin).
GN PKHD1 OR FCYT OR TICM1.
OS Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RN SEQUENCE FROM N.A. (ISOFORM 1), VARIANTS ARPKD MET-36; VAL-222;
RN TRP-1249; ARG-1407; PHE-1664; MET-1741; ARG-1917; GLY-1995; LYS-2331;
RN THR-2957; PHE-3018 AND THR-3553 AND VARIANTS MET-752; CYS-760;
RN ARG-852; VAL-1262; MET-2938; TYR-3139; ILE-3960 AND ARG-4048.
RN TISSUE=Kidney;
RX MEDLINE=21918597; PubMed=11919560;
Ward C.J., Hogan M.C., Rossetti S., Walker D., Sneddon T., Wang X.,
Kubly V., Cunningham J.M., Bacallao R., Ishibashi M., Milliner D.S.,
Torres V.E., Harris P.C.;
"The gene mutated in autosomal recessive polycystic kidney disease
encodes a large, receptor-like protein.";
Nat. Genet. 30:259-269(2002).
[2]
RN SEQUENCE FROM N.A. (ISOFORM 1), ALTERNATIVE SPLICING, AND VARIANTS
RN ARPKD MET-36; VAL-222; LEU-253; HIS-760; SER-1122; TRP-1624; THR-2957
RN AND TYR-3139.
RC TISSUE=Kidney;
RX MEDLINE=21947447; PubMed=11898128;
Onuchic L.F., Furu L., Nagasawa Y., Hou X., Eggemann T., Ren Z.,
Bergmann C., Senderek J., Esquivel E., Zeltner R.,
Rudnik-Schoenborn S., Mrug M., Sweeney W., Avner E.D., Zerres K.,
Guay-Woodford L.M., Somlo S., Germino G.G.;
"PKHD1, the polycystic kidney and hepatic disease 1 gene, encodes a
novel large protein containing multiple
immunoglobulin-like-plexin-transcription-factor domains and parallel
beta-helix 1 repeats";
Am. J. Hum. Genet. 70:1305-1317(2002).
[3]
RN SEQUENCE FROM N.A. (ISOFORM 2), AND VARIANT TYR-1204.
RC TISSUE=Petal kidney;
MEDLINE=22074934; PubMed=12079288;
Xiong H., Chen Y., Yi Y., Tsuchiya K., Moeckel G., Cheung J.,
Liang D., Tham K., Xu X., Chen X.-Z., Pei Y., Zhao Z.J., Wu G.;
"A novel gene encoding a TIG multiple domain protein is a positional
candidate for autosomal recessive polycystic kidney disease.";
Genomics 80:96-104(2002).
RL Genomics 80:96-104(2002).
CC -!- FUNCTION: May be a receptor protein that acts in collecting-duct
CC and biliary differentiation.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein (Probable).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Comment=Additional isoforms seem to exist. As a matter of fact,
CC alternatively spliced products seem to fall into two broad
CC groups: one group, which includes the longest continuous ORF but
CC which may also include molecules lacking some middle domains,
CC has a single TM element and is likely to be associated with the
CC plasma membrane. The other group lacks a TM domain and thus its
CC members may be secreted;
CC Name=1;
CC IsoId=Q8TCZ9-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q8TCZ9-2; Sequence=VSP_003947, VSP_003948;
CC -!- TISSUE SPECIFICITY: Predominantly expressed in fetal and adult
CC kidney. Also present in the adult pancreas, but at much lower
CC levels. Detectable in fetal and adult liver. Rather indistinct
CC signal in fetal brain.
CC -!- DISEASE: Defects in PKHD1 are the cause of all typical forms of
CC autosomal recessive polycystic kidney disease (ARPKD), a
CC hereditary and severe form of polycystic kidney disease affecting
CC the kidneys and the hepatic biliary tract. The clinical spectrum
CC is widely variable, with most cases presenting during infancy. The
CC fetal phenotypic features classically include enlarged and
CC echogenic kidneys, as well as oligohydramnios secondary to a poor
CC urine output. Up to 50% of the affected neonates die shortly after
CC birth, as a result of severe pulmonary hypoplasia and secondary
CC respiratory insufficiency. In the subset that survives the
CC perinatal period, morbidity and mortality are mainly related to
CC severe systemic hypertension, renal insufficiency, and portal
CC hypertension due to portal-tract fibrosis.
CC -!- SIMILARITY: Contains 9 Pih1 repeats.
CC -!- SIMILARITY: Contains 12 TIG/IFT domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
```


or send an email to license@sb-sib.ch).

CC
CC
DR EMBL; AY074797; AAL74290.1; -
DR EMBL; AY480064; AAM44232.1; -
DR EMBL; AY092083; AAM18186.1; -
DR Genew; HGNC:9016; PKHD1.
DR MIN; 606702; -
DR MIN; 263200; -
DR InterPro; IPR002909; IPT TIG.
DR InterPro; IPR006626; PBH1.
DR Pfam; PF01833; TIG; 7
DR SMART; SM00429; IPT; 6.
DR SMART; SM00710; PBH1; 10.
KW Signal; Glycoprotein; Transmembrane; Receptor; Repeat;
KW Alternative splicing; Disease mutation; Polymorphism.
FT SIGNAL 1 23
FT CHAIN 24 4074 POLYCYSTIC KIDNEY AND HEPATIC DISEASE 1.
FT DOMAIN 24 358 POTENTIAL.
FT TRANSMEM 3859 3879 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 3880 4074 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 24 111 TIG 1 (ATYPICAL).
FT DOMAIN 137 232 TIG 2.
FT DOMAIN 259 335 TIG 3.
FT DOMAIN 931 1015 TIG 4.
FT DOMAIN 1019 1103 TIG 5.
FT DOMAIN 1108 1192 TIG 6 (ATYPICAL).
FT DOMAIN 1196 1289 TIG 7.
FT DOMAIN 1301 1382 TIG 8 (ATYPICAL).
FT DOMAIN 1389 1481 TIG 9.
FT DOMAIN 1486 1570 TIG 10.
FT DOMAIN 1573 1599 TIG 11.
FT DOMAIN 1598 1658 TIG 12 (ATYPICAL).
FT REPEAT 2226 2248 PBH1 1.
FT REPEAT 2249 2271 PBH1 2.
FT REPEAT 2292 2325 PBH1 3.
FT REPEAT 2326 2347 PBH1 4.
FT REPEAT 2409 2431 PBH1 5.
FT REPEAT 2469 2502 PBH1 6.
FT REPEAT 3010 3032 PBH1 7.
FT REPEAT 3033 3055 PBH1 8.
FT REPEAT 3086 3108 PBH1 9.
FT CARBOHYD 54 54 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 226 226 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 276 276 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 357 357 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 387 387 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 507 507 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 520 520 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 529 529 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 641 641 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 711 711 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 830 830 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 869 869 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 967 967 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 977 977 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1065 1065 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1084 1084 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1116 1116 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1135 1135 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1234 1234 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1241 1241 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1241 1241 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1312 1312 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1323 1323 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1346 1346 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1377 1377 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1460 1460 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1475 1475 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1494 1494 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1527 1527 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1532 1532 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1564 1564 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1582 1582 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1602 1602 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1631 1631 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1698 1698 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1764 1764 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1779 1779 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1879 1879 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1883 1883 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1919 1919 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1945 1945 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1959 1959 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2034 2034 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2115 2115 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2144 2144 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2350 2350 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2385 2385 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2435 2435 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2471 2471 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2509 2509 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2535 2535 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2553 2553 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2595 2595 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2633 2633 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2753 2753 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2768 2768 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3008 3008 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3140 3140 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3169 3169 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3225 3225 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3488 3488 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3528 3528 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3707 3707 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3726 3726 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3838 3838 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 3386 3386 GTFREKQKCY -> VTTEQALKISE (in isoform 2).
FT VARSPLIC 3397 4074 /FTID=VSP_003947.
FT VARSPLIC 3397 4074 Missing (in isoform 2).
Query Match 49.4%; Score 44; DB 1; Length 4074;
Best Local Similarity 70.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 5 EDDFDWNPAD 14
Db 2065 EDAVDWNEGD 2074
ID LEQG BOTJR STANDARD; PRT; 135 AA.
RESULT 10
LEQG BOTJR
AC P83519; 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Galactose-specific lectin (HJCU).
OS Bothrops jararacussu (Jararacussu).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Viperidae; Crotalinae; Bothrops.
OX NCBI_TaxID=8726;
RN [1] SEQUENCE.
RP TISSUE=Venom;
RC MEDLINE=21899339; PubMed=11902666;
RA de Carvalho D.D., Marangoni S., Novello J.C.;
RT "Primary structure characterization of Bothrops jararacussu snake venom lectin."
RL [2] J. Protein Chem. 21:43-50(2002).
RN [2] SEQUENCE OF 1-8, FUNCTION, SUBUNIT, AND INDUCTION.
RP TISSUE=Venom;
RC de Carvalho D.D., Marangoni S., Oliveira B., Novello J.C.;
RA "Isolation and characterization of a new lectin from the venom of the snake Bothrops jararacussu."
RT

Biochem. Mol. Biol. Int. 44:933-938(1998).
 [3]
 RP FUNCTION.
 RX MEDLINE=20093835; PubMed=10628348;
 RA Pereira-Bittencourt M., de Carvalho D.D., Gagliardi A.R.,
 RA Collins D.C.;
 RT "The effect of a lectin from the venom of the snake, Bothrops
 jatacarussu, on tumor cell proliferation.";
 RL Anticancer Res. 19:4023-4025(1999).
 RN [4]
 RP FUNCTION.
 RX MEDLINE=21372655; PubMed=11478954;
 RA de Carvalho D.D., Schmitz S., Novello J.C., Markland F.S.;
 RA "Effect of BJGUL (a lectin from the venom of the snake Bothrops
 jatacarussu) on adhesion and growth of tumor and endothelial cells.";
 RL Toxicon 39:1471-1476(2001).
 CC -!- FUNCTION: Galactose-binding protein which recognizes specific
 CC carbohydrate structures and agglutinates a variety of animal cells
 CC by binding to cell-surface glycoproteins and glycolipids. Calcium-
 CC dependent lectin. Shows high hemagglutinating activity. An
 CC effective inhibitor of cell growth in some cancer cell lines,
 CC especially against renal and pancreatic cancer cell lines, human
 CC breast and ovarian carcinoma, glioblastoma and a bovine brain
 CC microvascular endothelial cell line.
 CC -!- SUBUNIT: Homodimer; disulfide-linked.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- INDUCTION: Hemagglutination activity is inhibited by lactose,
 CC galactose, raffinose, EDTA and EGTA.
 CC -!- SIMILARITY: Contains 1 C-type lectin family domain.
 DR GO: GO:0005576; F:galactose binding; TAS.
 DR GO: GO:000771; P:agglutination; TAS.
 DR GO: GO:0016338; P:calcium-dependent cell-cell adhesion; TAS.
 DR GO: GO:0030308; P:negative regulation of cell growth; TAS.
 DR InterPro: IPR001304; Lectin C.
 DR InterPro: IPR003990; Pancreatis_ac.
 DR Pfam: PF00059; lectin c; 1.
 DR PRINTS: PR01504; PNCREATTISAP.
 DR SMART: SM00034; CLECT; 1.
 DR SMART: PS00615; C TYPE LECTIN 1; 1.
 DR PROSITE: PS50041; C_TYPE_LECTIN_2; 1.
 KW Calcium; Lectin.
 FT DOMAIN 10 132 C-TYPE LECTIN.
 FT DISULFID 3 14 BY SIMILARITY.
 FT DISULFID 31 131 BY SIMILARITY.
 FT DISULFID 38 133 BY SIMILARITY.
 FT DISULFID 106 123 BY SIMILARITY.
 FT DISULFID 86 86 INTERCHAIN (BY SIMILARITY).
 SQ SEQUENCE 135 AA; 16113 MW; E1935CCD13AC69E CRC64;
 Query Match 48.3%; Score 43; DB 1; Length 135;
 Best Local Similarity 50.0%; Pred. No. 5.3;
 Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
 QY 2 QDREDFDWNPADR 15
 Db 71 QDKKDFSWETDR 84
 RESULT 11
 MD -MGN DROME STANDARD; PRT; 147 AA.
 AC P49028; Q9W2L3;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Mago nashi protein.
 GN MAGO OR MGN OR CG9401.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OC NCBI_taxid=7227;

[1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94298523; PubMed=8026338;
 RA Newmark P.A., Boswell R.E.;
 RT "The mago nashi locus encodes an essential product required for germ
 RL plasm assembly in Drosophila.";
 RN Development 120:1303-1313(1994).
 [2]
 RP SEQUENCE FROM N.A.
 RX STRAIN=Berkley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.C., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.V., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Fertiera S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei E., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J.J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusekern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 [3]
 RP FUNCTION.
 RX MEDLINE=97417587; PubMed=9272960;
 RA Newmark P.A., Mohr S.E., Gong L., Boswell R.E.;
 RT "Mago nashi mediates the posterior follicle cell-to-oocyte signal to
 RL organize axis formation in Drosophila.";
 RN Development 124:3197-3207(1997).
 [4]
 RP FUNCTION.
 RX MEDLINE=97362356; PubMed=9210377;
 RA Micklem D.R., Dasgupta R., Elliott H., Gergely F., Davidson C.,
 RA Brand A., Gonzalez-Reyes A., St Johnston D.;
 RT "The mago nashi gene is required for the polarisation of the oocyte
 RL and the formation of perpendicular axes in Drosophila.";
 CC CURR. BIOL. 7:468-478(1997).
 CC -!- FUNCTION: PARTICIPATES IN THE BIDIRECTIONAL INTERCELLULAR
 CC SIGNALING BETWEEN THE POSTERIOR FOLLICLE CELLS AND OOCYTE TO
 CC ESTABLISH SPATIAL COORDINATES THAT INDUCES AXIS FORMATION.
 CC REQUIRED FOR THE POLARIZATION OF THE OOCYTE MICROTUBULE
 CC CYTOSKELETON. MUTATIONS IN MGN DISRUPT THE STAUPEIN PROTEIN.
 CC LOCALIZATION OF THE OSKAR mRNA AND THE STAUPEIN PROTEIN.
 CC -!- SUBCELLULAR LOCATION: NUCLEAR; WITHIN THE POSTERIOR POLE PLASM.
 CC -!- MISCELLANEOUS: 'Mago nashi' means 'without grandchildren' in
 CC Japanese.

```

CC -!- SIMILARITY: Belongs to the mago nashi family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U03559; AAC13746.1; -.
CC EMBL; AE03453; AAF46677.1; -.
CC FLYBase; FBgn002736; mago.
CC DR GO; GO:0005737; C:cytoplasm; IDA.
CC DR GO; GO:0005634; C:nucleus; IDA.
CC DR GO; GO:0008103; P:occure microtubule cytoskeleton polarization; IMP.
CC DR GO; GO:0008104; P:protein localization; IMP.
CC DR GO; GO:0007317; P:regulation of pole plasm oskar mRNA localiz. .; IMP.
CC DR InterPro; IPR004023; Mago nashi.
CC DR Pfam; PF02792; Mago nashi; 1.
CC DR ProDom; PD009481; Mago nashi; 1.
CC KW Developmental protein; Nuclear protein.
CC FT VARIANT 19 19 G -> R (IN ALLELE MAGO-1).
CC FT VARIANT 91 91 I -> T (IN ALLELE MAGO-WF7).
CC SQ SEQUENCE 147 AA; 17311 MW; B9A2CBFC05EBEFD CRC64;
CC -----
Query Match 48.3%; Score 43; DB 1; Length 147;
Best Local Similarity 58.3%; Pred. No. 5.8; Indels 0; Gaps 0;
Matches 7; Conservative 1; Mismatches 4;
QY 4 REDDEDWNPADR 15
DB 72 QEDDLPPPPDR 83
RESULT 12
FCP1 SCHPO STANDARD; PRT; 723 AA.
AC Q9P376;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phosphoserine phosphatase fcp1 (EC 3.1.3.3) (CTD phosphatase fcp1).
GN FCP1 OR SPAC19B12.05C.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgourou J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby I., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,
RA Weltjens J., Volkart G., Aert R., Robben J., Grymonprez B.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,

```

```

RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
RN [2]
RP FUNCTION, COFACTOR, SUBUNIT, CATALYTIC ACTIVITY, ACTIVE SITE, AND
RP MUTAGENESIS.
RX MEDLINE=22050624; PubMed=11934898;
RA Hausmann S., Shuman S.;
RT "Characterization of the CTD phosphatase Fcp1 from fission yeast.
RT Preferential dephosphorylation of serine 2 versus serine 5.";
RL J. Biol. Chem. 277:21213-21220(2002).
CC -!- FUNCTION: Promotes activity of RNA polymerase II elongation factor
CC where it dephosphorylates Ser-2 and Ser-5.
CC -!- CATALYTIC ACTIVITY: Phosphoserine + H(2)O = serine + phosphate.
CC -!- COFACTOR: Magnesium or manganese or cobalt.
CC -!- SUBUNIT: Monomer.
CC -!- SIMILARITY: Contains 1 BRCT domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AL390814; CAC00553.1; -.
CC GeneDB SPombe; SPAC19B12.05C; -.
CC InterPro; IPR001357; BRCT.
CC InterPro; IPR004274; NIF.
CC Pfam; PF00533; BRCT; 1.
CC Pfam; PF03031; NIF; 1.
CC SMART; SM00292; BRCT; 1.
CC SMART; SM00577; CPDC; 1.
CC PROSITE; PS0172; BRCT; 1.
CC Hydrolase; Metal-binding; Magnesium; Manganese; Cobalt.
KW DOMAIN 140 580 PHOSPHATASE.
FT DOMAIN 486 579 BRCT.
FT ACT_SITE 170 170
FT ACT_SITE 172 172
FT MUTAGEN 170 170 D->A,N,E: NO ACTIVITY.
FT MUTAGEN 172 172 D->A,N,E: NO ACTIVITY.
SQ SEQUENCE 723 AA; 81964 MW; A127A06CD2FD3435 CRC64;
Query Match 48.3%; Score 43; DB 1; Length 723;
Best Local Similarity 70.0%; Pred. No. 33;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 3 DREDDFWNP 12
DB 298 DRGDVWDWNP 307
RESULT 13
PHFE MOUSE
ID PHFE MOUSE STANDARD; PRT; 881 AA.
AC Q9DAH9; Q81026; Q8CGF8;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE PHD finger protein 14.
DE PHD finger protein 14.
OS Mus musculus (Mouse).
GN PHF14.
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=21085660; PubMed=11217851;

```

Fri May 21 12:56:11 2004

```

RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Azawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikiado I., Peole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tonita M., Wagner L., Barsh G.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Washima J., Mazzarelli J., Momberto N.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N;
RX MEDLINE=22388257; PubMed=12477932;
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Locquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Rosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalus D.E.,
RA Schermer A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RP SEQUENCE OF 513-881 FROM N.A.
RC STRAIN=129S6/SvEvTac;
RA Brathwaite M., Waelitz P., Schlessinger D., Nagaraja R.;
RT "Genomic sequence analysis in the mouse T-complex region.";
RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Contains 2 PHD-type zinc fingers.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; AK016517; BAB30282.1; -
CC DR EMBL; BC040236; AAH40236.1; -
CC DR EMBL; AY162410; AA017166.1; -
CC DR MGD; MGI:1922975; 4932409F11rlk.
CC DR InterPro; IPR001965; Znf_PHD.
CC DR Pfam; PF00628; PHD; 2.
CC DR SMART; SM00249; PHD; 3.
CC DR PROSITE; PS01359; ZF_PHD_1; 2.
CC DR PROSITE; PS00016; ZF_PHD_2; 2.
CC DR Zinc-finger; Repeat.
CC KW DOMAIN 62 129 GLU-RICH.
CC FT ZN_FING 312 373 PHD-TYPE 1.
CC FT ZN_FING 718 772 PHD-TYPE 2.
CC
CC -----
CC EMBL; AK016517; BAB30282.1; -
CC DR EMBL; BC040236; AAH40236.1; -
CC DR EMBL; AY162410; AA017166.1; -
CC DR MGD; MGI:1922975; 4932409F11rlk.
CC DR InterPro; IPR001965; Znf_PHD.
CC DR Pfam; PF00628; PHD; 2.
CC DR SMART; SM00249; PHD; 3.
CC DR PROSITE; PS01359; ZF_PHD_1; 2.
CC DR PROSITE; PS00016; ZF_PHD_2; 2.
CC DR Zinc-finger; Repeat.
CC KW DOMAIN 62 129 GLU-RICH.
CC FT ZN_FING 312 373 PHD-TYPE 1.
CC FT ZN_FING 718 772 PHD-TYPE 2.
CC
CC -----
CC EMBL; AK016517; BAB30282.1; -
CC DR EMBL; BC040236; AAH40236.1; -
CC DR EMBL; AY162410; AA017166.1; -
CC DR MGD; MGI:1922975; 4932409F11rlk.
CC DR InterPro; IPR001965; Znf_PHD.
CC DR Pfam; PF00628; PHD; 2.
CC DR SMART; SM00249; PHD; 3.
CC DR PROSITE; PS01359; ZF_PHD_1; 2.
CC DR PROSITE; PS00016; ZF_PHD_2; 2.
CC DR Zinc-finger; Repeat.
CC KW DOMAIN 62 129 GLU-RICH.
CC FT ZN_FING 312 373 PHD-TYPE 1.
CC FT ZN_FING 718 772 PHD-TYPE 2.
CC
CC -----
CC EMBL; AK016517; BAB30282.1; -
CC DR EMBL; BC040236; AAH40236.1; -
CC DR EMBL; AY162410; AA017166.1; -
CC DR MGD; MGI:1922975; 4932409F11rlk.
CC DR InterPro; IPR001965; Znf_PHD.
CC DR Pfam; PF00628; PHD; 2.
CC DR SMART; SM00249; PHD; 3.
CC DR PROSITE; PS01359; ZF_PHD_1; 2.
CC DR PROSITE; PS00016; ZF_PHD_2; 2.
CC DR Zinc-finger; Repeat.
CC KW DOMAIN 62 129 GLU-RICH.
CC FT ZN_FING 312 373 PHD-TYPE 1.
CC FT ZN_FING 718 772 PHD-TYPE 2.
CC
CC -----
CC EMBL; AK016517; BAB30282.1; -
CC DR EMBL; BC040236; AAH40236.1; -
CC DR EMBL; AY162410; AA017166.1; -
CC DR MGD; MGI:1922975; 4932409F11rlk.
CC DR InterPro; IPR001965; Znf_PHD.
CC DR Pfam; PF00628; PHD; 2.
CC DR SMART; SM00249; PHD; 3.
CC DR PROSITE; PS01359; ZF_PHD_1; 2.
CC DR PROSITE; PS00016; ZF_PHD_2; 2.
CC DR Zinc-finger; Repeat.
CC KW DOMAIN 62 129 GLU-RICH.
CC FT ZN_FING 312 373 PHD-TYPE 1.
CC FT ZN_FING 718 772 PHD-TYPE 2.
CC
CC -----
CC EMBL; AK016517; BAB30282.1; -
CC DR EMBL; BC040236; AAH40236.1; -
CC DR EMBL; AY162410; AA017166.1; -
CC DR MGD; MGI:1922975; 4932409F11rlk.
CC DR InterPro; IPR001965; Znf_PHD.
CC DR Pfam; PF00628; PHD; 2.
CC DR SMART; SM00249; PHD; 3.
CC DR PROSITE; PS01359; ZF_PHD_1; 2.
CC DR PROSITE; PS00016; ZF_PHD_2; 2.
CC DR Zinc-finger; Repeat.
CC KW DOMAIN 62 129 GLU-RICH.
CC FT ZN_FING 312 373 PHD-TYPE 1.
CC FT ZN_FING 718 772 PHD-TYPE 2.
CC
CC -----
CC EMBL; AK016517; BAB30282.1; -
CC DR EMBL; BC040236; AAH40236.1; -
CC DR EMBL; AY162410; AA017166.1; -
CC DR MGD; MGI:1922975; 4932409F11rlk.
CC DR InterPro; IPR001965; Znf_PHD.
CC DR Pfam; PF00628; PHD; 2.
CC DR SMART; SM00249; PHD; 3.
CC DR PROSITE; PS01359; ZF_PHD_1; 2.
CC DR PROSITE; PS00016; ZF_PHD_2; 2.
CC DR Zinc-finger; Repeat.
CC KW DOMAIN 62 129 GLU-RICH.
CC FT ZN_FING 312 373 PHD-TYPE 1.
CC FT ZN_FING 718 772 PHD-TYPE 2.
CC
CC -----
CC EMBL; AK016517; BAB30282.1; -
CC DR EMBL; BC040236; AAH40236.1; -
CC DR EMBL; AY162410; AA017166.1; -
CC DR MGD; MGI:1922975; 4932409F11rlk.
CC DR InterPro; IPR001965; Znf_PHD.
CC DR Pfam; PF00628; PHD; 2.
CC DR SMART; SM00249; PHD; 3.
CC DR PROSITE; PS01359; ZF_PHD_1; 2.
CC DR PROSITE; PS00016; ZF_PHD_2; 2.
CC DR Zinc-finger; Repeat.
CC KW DOMAIN 62 129 GLU-RICH.
CC FT ZN_FING 312 373 PHD-TYPE 1.
CC FT ZN_FING 718 772 PHD-TYPE 2.
CC
CC -----
CC EMBL; AK016517; BAB30282.1; -
CC DR EMBL; BC040236; AAH40236.1; -
CC DR EMBL; AY162410; AA017166.1; -
CC DR MGD; MGI:1922975; 4932409F11rlk.
CC DR InterPro; IPR001965; Znf_PHD.
CC DR Pfam; PF00628; PHD; 2.
CC DR SMART; SM00249; PHD; 3.
CC DR PROSITE; PS01359; ZF_PHD_1; 2.
CC DR PROSITE; PS00016; ZF_PHD_2; 2.
CC DR Zinc-finger; Repeat.
CC KW DOMAIN 62 129 GLU-RICH.
CC FT ZN_FING 312 373 PHD-TYPE 1.
CC FT ZN_FING 718 772 PHD-TYPE 2.
CC
CC -----
CC EMBL; AK016517; BAB30282.1; -
CC DR EMBL; BC040236; AAH40236.1; -
CC DR EMBL; AY162410; AA017166.1; -
CC DR MGD; MGI:1922975; 4932409F11rlk.
CC DR InterPro; IPR001965; Znf_PHD.
CC DR Pfam; PF00628; PHD; 2.
CC DR SMART; SM00249; PHD; 3.
CC DR PROSITE; PS01359; ZF_PHD_1; 2.
CC DR PROSITE; PS00016; ZF_PHD_2; 2.
CC DR Zinc-finger; Repeat.
CC KW DOMAIN 62 129 GLU-RICH.
CC FT ZN_FING 312 373 PHD-TYPE 1.
CC FT ZN_FING 718 772 PHD-TYPE 2.
CC
CC -----
CC EMBL; AK016517; BAB30282.1; -
CC DR EMBL; BC040236; AAH40236.1; -
CC DR EMBL; AY162410; AA017166.1; -
CC DR MGD; MGI:1922975; 4932409F11rlk.
CC DR InterPro; IPR001965; Znf_PHD.
CC DR Pfam; PF00628; PHD; 2.
CC DR SMART; SM00249; PHD; 3.
CC DR PROSITE; PS01359; ZF_PHD_1; 2.
CC DR PROSITE; PS00016; ZF_PHD_2; 2.
CC DR Zinc-finger; Repeat.
CC KW DOMAIN 62 129 GLU-RICH.
CC FT ZN_FING 312 373 PHD-TYPE 1.
CC FT ZN_FING 718 772 PHD-TYPE 2.
CC
CC -----
CC EMBL; AK016517; BAB30282.1; -
CC DR EMBL; BC040236; AAH40236.1; -
CC DR EMBL; AY162410; AA017166.1; -
CC DR MGD; MGI:1922975; 4932409F11rlk.
CC DR InterPro; IPR001965; Znf_PHD.
CC DR Pfam; PF00628; PHD; 2.
CC DR SMART; SM00249; PHD; 3.
CC DR PROSITE; PS01359; ZF_PHD_1; 2.
CC DR PROSITE; PS00016; ZF_PHD_2; 2.
CC DR Zinc-finger; Repeat.
CC KW DOMAIN 62 129 GLU-RICH.
CC FT ZN_FING 312 373 PHD-TYPE 1.
CC FT ZN_FING 718 772 PHD-TYPE 2.
CC
CC -----
CC EMBL; AK016517; BAB30282.1; -
CC DR EMBL; BC040236; AAH40236.1; -
CC DR EMBL; AY162410; AA017166.1; -
CC DR MGD; MGI:1922975; 4932409F11rlk.
CC DR InterPro; IPR001965; Znf_PHD.
CC DR Pfam; PF00628; PHD; 2.
CC DR SMART; SM00249; PHD; 3.
CC DR PROSITE; PS01359; ZF_PHD_1; 2.
CC DR PROSITE; PS00016; ZF_PHD_2; 2.
CC DR Zinc-finger; Repeat.
CC KW DOMAIN 62 129 GLU-RICH.
CC FT ZN_FING 312 373 PHD-TYPE 1.
CC FT ZN_FING 718 772 PHD-TYPE 2.
CC
CC -----
CC EMBL; AK016517; BAB30282.1; -
CC DR EMBL; BC040236; AAH40236.1; -
CC DR EMBL; AY162410; AA017166.1; -
CC DR MGD; MGI:1922975; 4932409F11rlk.
CC DR InterPro; IPR001965; Znf_PHD.
CC DR Pfam; PF00628; PHD; 2.
CC DR SMART; SM00249; PHD; 3.
CC DR PROSITE; PS01359; ZF_PHD_1; 2.
CC DR PROSITE; PS00016; ZF_PHD_2; 2.
CC DR Zinc-finger; Repeat.
CC KW DOMAIN 62 129 GLU-RICH.
CC FT ZN_FING 312 373 PHD-TYPE 1.
CC FT ZN_FING 718 772 PHD-TYPE 2.
CC
CC -----
CC EMBL; AK016517; BAB30282.1; -
CC DR EMBL; BC040236; AAH40236.1; -
CC DR EMBL; AY162410; AA017166.1; -
CC DR MGD; MGI:1922975; 4932409F11rlk.
CC DR InterPro; IPR001965; Znf_PHD.
CC DR Pfam; PF00628; PHD; 2.
CC DR SMART; SM00249; PHD; 3.
CC DR PROSITE; PS01359; ZF_PHD_1; 2.
CC DR PROSITE; PS00016; ZF_PHD_2; 2.
CC DR Zinc-finger; Repeat.
CC KW DOMAIN 62 129 GLU-RICH.
CC FT ZN_FING 312 373 PHD-TYPE 1.
CC FT ZN_FING 718 772 PHD-TYPE 2.
CC
CC -----
CC EMBL; AK016517; BAB30282.1; -
CC DR EMBL; BC040236; AAH40236.1; -
CC DR EMBL; AY162410; AA017166.1; -
CC DR MGD; MGI:1922975; 4932409F11rlk.
CC DR InterPro; IPR001965; Znf_PHD.
CC DR Pfam; PF00628; PHD; 2.
CC DR SMART; SM00249; PHD; 3.
CC DR PROSITE; PS01359; ZF_PHD_1; 2.
CC DR PROSITE; PS00016; ZF_PHD_2; 2.
CC DR Zinc-finger; Repeat.
CC KW DOMAIN 62 129 GLU-RICH.
CC FT ZN_FING 312 373 PHD-TYPE 1.
CC FT ZN_FING 718 772 PHD-TYPE 2.
CC
CC -----
CC EMBL; AK016517; BAB30282.1; -
CC DR EMBL; BC040236; AAH40236.1; -
CC DR EMBL; AY162410; AA017166.1; -
CC DR MGD; MGI:1922975; 4932409F11rlk.
CC DR InterPro; IPR001965; Znf_PHD.
CC DR Pfam; PF00628; PHD; 2.
CC DR SMART; SM00249; PHD; 3.
CC DR PROSITE; PS01359; ZF_PHD_1; 2.
CC DR PROSITE; PS00016; ZF_PHD_2; 2.
CC DR Zinc-finger; Repeat.
CC KW DOMAIN 62 129 GLU-RICH.
CC FT ZN_FING 312 373 PHD-TYPE 1.
CC FT ZN_FING 718 772 PHD-TYPE 2.
CC
CC -----
CC EMBL; AK016517; BAB30282.1; -
CC DR EMBL; BC040236; AAH40236.1; -
CC DR EMBL; AY162410; AA017166.1; -
CC DR MGD; MGI:1922975; 4932409F11rlk.
CC DR InterPro; IPR001965; Znf_PHD.
CC DR Pfam; PF00628; PHD; 2.
CC DR SMART; SM00249; PHD; 3.
CC DR PROSITE; PS01359; ZF_PHD_1; 2.
CC DR PROSITE; PS00016; ZF_PHD_2; 2.
CC DR Zinc-finger; Repeat.
CC KW DOMAIN 62 129 GLU-RICH.
CC FT ZN_FING 312 373 PHD-TYPE 1.
CC FT ZN_FING 718 772 PHD-TYPE 2.
CC
CC -----
CC EMBL; AK016517; BAB30282.1; -
CC DR EMBL; BC040236; AAH40236.1; -
CC DR EMBL; AY162410; AA017166.1; -
CC DR MGD; MGI:1922975; 4932409F11rlk.
CC DR InterPro; IPR001965; Znf_PHD.
CC DR Pfam; PF00628; PHD; 2.
CC DR SMART; SM00249; PHD; 3.
CC DR PROSITE; PS01359; ZF_PHD_1; 2.
CC DR PROSITE; PS00016; ZF_PHD_2; 2.
CC DR Zinc-finger; Repeat.
CC KW DOMAIN 62 129 GLU-RICH.
CC FT ZN_FING 312 373 PHD-TYPE 1.
CC FT ZN_FING 718 772 PHD-TYPE 2.
CC
CC -----
CC EMBL; AK016517; BAB30282.1; -
CC DR EMBL; BC040236; AAH40236.1; -
CC DR EMBL; AY162410; AA017166.1; -
CC DR MGD; MGI:1922975; 4932409F11rlk.
CC DR InterPro; IPR001965; Znf_PHD.
CC DR Pfam; PF00628; PHD; 2.
CC DR SMART; SM00249; PHD; 3.
CC DR PROSITE; PS01359; ZF_PHD_1; 2.
CC DR PROSITE; PS00016; ZF_PHD_2; 2.
CC DR Zinc-finger; Repeat.
CC KW DOMAIN 62 129 GLU-RICH.
CC FT ZN_FING 312 373 PHD-TYPE 1.
CC FT ZN_FING 718 772 PHD-TYPE 2.
CC
CC -----
CC EMBL; AK016517; BAB30282.1; -
CC DR EMBL; BC040236; AAH40236.1; -
CC DR EMBL; AY162410; AA017166.1; -
CC DR MGD; MGI:1922975; 4932409F11rlk.
CC DR InterPro; IPR001965; Znf_PHD.
CC DR Pfam; PF00628; PHD; 2.
CC DR SMART; SM00249; PHD; 3.
CC DR PROSITE; PS01359; ZF_PHD_1; 2.
CC DR PROSITE; PS00016; ZF_PHD_2; 2.
CC DR Zinc-finger; Repeat.
CC KW DOMAIN 62 129 GLU-RICH.
CC FT ZN_FING 312 373 PHD-TYPE 1.
CC FT ZN_FING 718 772 PHD-TYPE 2.
CC
CC -----
CC EMBL; AK016517; BAB30282.1; -
CC DR EMBL; BC040236; AAH40236.1; -
CC DR EMBL; AY162410; AA017166.1; -
CC DR MGD; MGI:1922975; 4932409F11rlk.
CC DR InterPro; IPR001965; Znf_PHD.
CC DR Pfam; PF00628; PHD; 2.
CC DR SMART; SM00249; PHD; 3.
CC DR PROSITE; PS01359; ZF_PHD_1; 2.
CC DR PROSITE; PS00016; ZF_PHD_2; 2.
CC DR Zinc-finger; Repeat.
CC KW DOMAIN 62 129 GLU-RICH.
CC FT ZN_FING 312 373 PHD-TYPE 1.
CC FT ZN_FING 718 772 PHD-TYPE 2.
CC
CC -----
CC EMBL; AK016517; BAB30282.1; -
CC DR EMBL; BC040236; AAH40236.1; -
CC DR EMBL; AY162410; AA017166.1; -
CC DR MGD; MGI:1922975; 4932409F11rlk.
CC DR InterPro; IPR001965; Znf_PHD.
CC DR Pfam; PF00628; PHD; 2.
CC DR SMART; SM00249; PHD; 3.
CC DR PROSITE; PS01359; ZF_PHD_1; 2.
CC DR PROSITE; PS00016; ZF_PHD_2; 2.
CC DR Zinc-finger; Repeat.
CC KW DOMAIN 62 129 GLU-RICH.
CC FT ZN_FING 312 373 PHD-TYPE 1.
CC FT ZN_FING 718 772 PHD-TYPE 2.
CC
CC -----
CC EMBL; AK016517; BAB30282.1; -
CC DR EMBL; BC040236; AAH40236.1; -
CC DR EMBL; AY162410; AA017166.1; -
CC DR MGD; MGI:1922975; 4932409F11rlk.
CC DR InterPro; IPR001965; Znf_PHD.
CC DR Pfam; PF00628; PHD; 2.
CC DR SMART; SM00249; PHD; 3.
CC DR PROSITE; PS01359; ZF_PHD_1; 2.
CC DR PROSITE; PS00016; ZF_PHD_2; 2.
CC DR Zinc-finger; Repeat.
CC KW DOMAIN 62 129 GLU-RICH.
CC FT ZN_FING 312 373 PHD-TYPE 1.
CC FT ZN_FING 718 772 PHD-TYPE 2.
CC
CC -----
CC EMBL; AK016517; BAB30282.1; -
CC DR EMBL; BC040236; AAH40236.1; -
CC DR EMBL; AY162410; AA017166.1; -
CC DR MGD; MGI:1922975; 4932409F11rlk.
CC DR InterPro; IPR001965; Znf_PHD.
CC DR Pfam; PF00628; PHD; 2.
CC DR SMART; SM00249; PHD; 3.
CC DR PROSITE; PS01359; ZF_PHD_1; 2.
CC DR PROSITE; PS00016; ZF_PHD_2; 2.
CC DR Zinc-finger; Repeat.
CC KW DOMAIN 62 129 GLU-RICH.
CC FT ZN_FING 312 373 PHD-TYPE 1.
CC FT ZN_FING 718 772 PHD-TYPE 2.
CC
CC -----
CC EMBL; AK016517; BAB30282.1; -
CC DR EMBL; BC040236; AAH40236.1; -
CC DR EMBL; AY162410; AA017166.1; -
CC DR MGD; MGI:1922975; 4932409F11rlk.
CC DR InterPro; IPR001965; Znf_PHD.
CC DR Pfam; PF00628; PHD; 2.
CC DR SMART; SM00249; PHD; 3.
CC DR PROSITE; PS01359; ZF_PHD_1; 2.
CC DR PROSITE; PS00016; ZF_PHD_2; 2.
CC DR Zinc-finger; Repeat.
CC KW DOMAIN 62 129 GLU-RICH.
CC FT ZN_FING 312 373 PHD-TYPE 1.
CC FT ZN_FING 718 772 PHD-TYPE 2.
CC
CC -----
CC EMBL; AK016517; BAB30282.1; -
CC DR EMBL; BC040236; AAH40236.1; -
CC DR EMBL; AY162410; AA017166.1; -
CC DR MGD; MGI:1922975; 4932409F11rlk.
CC DR InterPro; IPR001965; Znf_PHD.
CC DR Pfam; PF00628; PHD; 2.
CC DR SMART; SM00249; PHD; 3.
CC DR PROSITE; PS01359; ZF_PHD_1; 2.
CC DR PROSITE; PS00016; ZF_PHD_2; 2.
CC DR Zinc-finger; Repeat.
CC KW DOMAIN 62 129 GLU-RICH.
CC FT ZN_FING 312 373 PHD-TYPE 1.
CC FT ZN_FING 718 772 PHD-TYPE 2.
CC
CC -----
CC EMBL; AK016517; BAB30282.1; -
CC DR EMBL; BC040236; AAH40236.1; -
CC DR EMBL; AY162410; AA017166.1; -
CC DR MGD; MGI:1922975; 4932409F11rlk.
CC DR InterPro; IPR001965; Znf_PHD.
CC DR Pfam; PF00628; PHD; 2.
CC DR SMART; SM00249; PHD; 3.
CC DR PROSITE; PS01359; ZF_PHD_1; 2.
CC DR PROSITE; PS00016; ZF_PHD_2; 2.
CC DR Zinc-finger; Repeat.
CC KW DOMAIN 62 129 GLU-RICH.
CC FT ZN_FING 312 373 PHD-TYPE 1.
CC FT ZN_FING 718 772 PHD-TYPE 2.
CC
CC -----
CC EMBL; AK016517; BAB30282.1; -
CC DR EMBL; BC040236; AAH40236.1; -
CC DR EMBL; AY162410; AA017166.1; -
CC DR MGD; MGI:1922975; 4932409F11rlk.
CC DR InterPro; IPR001965; Znf_PHD.
CC DR Pfam; PF00628; PHD; 2.
CC DR SMART; SM00249; PHD; 3.
CC DR PROSITE; PS01359; ZF_PHD_1; 2.
CC DR PROSITE; PS00016; ZF_PHD_2; 2.
CC DR Zinc-finger; Repeat.
CC KW DOMAIN 62 129 GLU-RICH.
CC FT ZN_FING 312 373 PHD-TYPE 1.
CC FT ZN_FING 718 772 PHD-TYPE 2.
CC
CC -----
CC EMBL; AK016517; BAB30282.1; -
CC DR EMBL; BC040236; AAH40236.1; -
CC DR EMBL; AY162410; AA017166.1; -
CC DR MGD; MGI:1922975; 4932409F11rlk.
CC DR InterPro; IPR001965; Znf_PHD.
CC DR Pfam; PF00628; PHD; 2.
CC DR SMART; SM00249; PHD; 3.
CC DR PROSITE; PS01359; ZF_PHD_1; 2.
CC DR PROSITE; PS00016; ZF_PHD_2; 2.
CC DR Zinc-finger; Repeat.
CC KW DOMAIN 62 129 GLU-RICH.
CC FT ZN_FING 312 373 PHD-TYPE 1.
CC FT ZN_FING 718 772 PHD-TYPE 2.
CC
CC -----
CC EMBL; AK016517; BAB30282.1; -
CC DR EMBL; BC040236; AAH40236.1; -
CC DR EMBL; AY162410; AA017166.1; -
CC DR MGD; MGI:1922975; 4932409F11rlk.
CC DR InterPro; IPR001965; Znf_PHD.
CC DR Pfam; PF00628; PHD; 2.
CC DR SMART; SM00249; PHD; 3.
CC DR PROSITE; PS01359; ZF_PHD_1; 2.
CC DR PROSITE; PS00016; ZF_PHD_2; 2.
CC DR Zinc-finger; Repeat.
CC KW DOMAIN 62 129 GLU-RICH.
CC FT ZN_FING 312 373 PHD-TYPE 1.
CC FT ZN_FING 718 772 PHD-TYPE 2.
CC
CC -----
CC EMBL; AK016517; BAB30282.1; -
CC DR EMBL; BC040236; AAH40236.1; -
CC DR EMBL; AY162410; AA017166.1; -
CC DR MGD; MGI:1922975; 4932409F11rlk.
CC DR InterPro; IPR001965; Znf_PHD.
CC DR Pfam; PF00628; PHD; 2.
CC DR SMART; SM00249; PHD; 3.
CC DR PROSITE; PS01359; ZF_PHD_1; 2.
CC DR PROSITE; PS00016; ZF_PHD_2; 2.
CC DR Zinc-finger; Repeat.
CC KW DOMAIN 62 129 GLU-RICH.
CC FT ZN_FING 312 373 PHD-TYPE 1.
CC FT ZN_FING 718 772 PHD-TYPE 2.
CC
CC -----
CC EMBL; AK016517; BAB30282.1; -
CC DR EMBL; BC040236; AAH40236.1; -
CC DR EMBL; AY162410; AA017166.1; -
CC DR MGD; MGI:1922975; 4932409F11rlk.
CC DR InterPro; IPR001965; Znf_PHD.
CC DR Pfam; PF00628; PHD; 2.
CC DR SMART; SM00249; PHD; 3.
CC DR PROSITE; PS01359; ZF_PHD_1; 2.
CC DR PROSITE; PS00016; ZF_PHD_2; 2.
CC DR Zinc-finger; Repeat.
CC KW DOMAIN 62 129 GLU-RICH.
CC FT ZN_FING 312 373 PHD-TYPE 1.
CC FT ZN_FING 718 772 PHD-TYPE 2.
CC
CC -----
CC EMBL; AK016517; BAB30282.1; -
CC DR EMBL; BC040236; AAH40236.1; -
CC DR EMBL; AY162410; AA017166.1; -
CC DR MGD; MGI:1922975; 4932409F11rlk.
CC DR InterPro; IPR001965; Znf_PHD.
CC DR Pfam; PF00628; PHD; 2.
CC DR SMART; SM00249; PHD; 3.
CC DR PROSITE; PS01359; ZF_PHD_1; 2.
CC DR PROSITE; PS00016; ZF_PHD_2; 2.
CC DR Zinc-finger; Repeat.
CC KW DOMAIN 62 129 GLU-RICH.
CC FT ZN_FING 312 373 PHD-TYPE 1.
CC FT ZN_FING 718 772 PHD-TYPE 2.
CC
CC -----
CC EMBL; AK016517; BAB30282.1; -
CC DR EMBL; BC040236; AAH40236.1; -
CC DR EMBL; AY162410; AA017166.1; -
CC DR MGD; MGI:1922975; 4932409F11rlk.
CC DR InterPro; IPR001965; Znf_PHD.
CC DR Pfam; PF00628; PHD; 2.
CC DR SMART; SM00249; PHD; 3.
CC DR PROSITE; PS01359; ZF_PHD_1; 2.
CC DR PROSITE; PS00016; ZF_PHD_2; 2.
CC DR Zinc-finger; Repeat.
CC KW DOMAIN 62 129 GLU-RICH.
CC FT ZN_FING 312 373 PHD-TYPE 1.
CC FT ZN_FING 718 772 PHD-TYPE 2.
CC
CC -----
CC EMBL; AK016517; BAB30282.1; -
CC DR EMBL; BC040236; AAH40236.1; -
CC DR EMBL; AY162410; AA017166.1; -
CC DR MGD; MGI:1922975; 4932409F11rlk.
CC DR InterPro; IPR001965; Znf_PHD.
CC DR Pfam; PF00628; PHD; 2.
CC DR SMART; SM00249; PHD; 3.
CC DR PROSITE; PS01359; ZF_PHD_1; 2.
CC DR PROSITE; PS00016; ZF_PHD_2; 2.
CC DR Zinc-finger; Repeat.
CC KW DOMAIN 62 129 GLU-RICH.
CC FT ZN_FING 312 373 PHD-TYPE 1.
CC FT ZN_FING 718 772 PHD-TYPE 2.
CC
CC -----
CC EMBL; AK016517; BAB30282.1; -
CC DR EMBL; BC040236; AAH40236.1; -
CC DR EMBL; AY162410; AA017166.1; -
CC DR MGD; MGI:1922975; 4932409F11rlk.
CC DR InterPro; IPR001965; Znf_PHD.
CC DR Pfam; PF00628; PHD; 2.
CC DR SMART; SM00249; PHD; 3.
CC DR PROSITE; PS01359; ZF_PHD_1; 2.
CC DR PROSITE; PS00016; ZF_PHD_2; 2.
CC DR Zinc-finger; Repeat.
CC KW DOMAIN 62 129 GLU-RICH.
CC FT ZN_FING 312 373 PHD-TYPE 1.
CC FT ZN_FING 718 772 PHD-TYPE 2.
CC
CC -----
CC EMBL; AK016517; BAB30282.1; -
CC DR EMBL; BC040236; AAH40236.1; -
CC DR EMBL; AY162410; AA017166.1; -
CC DR MGD; MGI:1922975; 4932409F11rlk.
CC DR InterPro; IPR001965; Znf_PHD.
CC DR Pfam; PF00628; PHD; 2.
CC DR SMART; SM00249; PHD; 3.
CC DR PROSITE; PS01359; ZF_PHD_1; 2.
CC DR PROSITE; PS00016; ZF_PHD_2; 2.
CC DR Zinc-finger; Repeat.
CC KW DOMAIN 62 129 GLU-RICH.
CC FT ZN_FING 312 373 PHD-TYPE 1.
CC FT ZN_FING 718 772 PHD-TYPE 2.
CC
CC -----
CC EMBL; AK016517; BAB30282.1; -
CC DR EMBL; BC040236; AAH40236.1; -
CC DR EMBL; AY162410; AA017166.1; -
CC DR MGD; MGI:1922975; 4932409F11rlk.
CC DR InterPro; IPR001965; Znf_PHD.
CC DR Pfam; PF00628; PHD; 2.
CC DR SMART; SM00249; PHD; 3.
CC DR PROSITE; PS01359; ZF_PHD_1; 2.
CC DR PROSITE; PS00016; ZF_PHD_2; 
```

Fri May 21 12:56:11 2004

```

AC Q9ZEA8;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable O-sialoglycoprotein endopeptidase (EC 3.4.24.57)
DE (Glycoprotease).
GN GCP OR RP037.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=782;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Madrid E;
RX MEDLINE=99039499; PubMed=9823893;
RA Andersson S.G.E., Zomorodipour A., Andersson J.O., Naeslund A.K.,
RA Sacheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
RT mitochondria."
RL Nature 396:133-140(1998).
CC -!- FUNCTION: Could be a metalloprotease.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of O-sialoglycoproteins; cleaves
CC 31-Arg-|-Asp-32 bond in glycoprotein A. Does not cleave
CC unglycosylated proteins, desialylated glycoproteins or
CC glycoproteins that are only N-glycosylated.
CC -!- COFACTOR: Zinc (Probable).
CC -!- SIMILARITY: Belongs to peptidase family M22.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ235270; CAA14508.1; -
CC PIR; E71711; E71711.
CC MEROPS; M22.UFW; -.
CC InterPro; IPR009005; Peptidase_M22.
CC Pfam; PF00814; Peptidase_M22; 1.
CC PIRSF; PIRSF004537; Osialgic_pttds; 1.
CC PRINTS; PR00789; OSIALOPTASE.
CC ProDom; PD02367; Peptidase_M22; 2.
CC TIGRFAMs; TIGR00329; GCP; 1.
CC PROSITE; PS01016; GLYCOPROTEASE; 1.
CC Hydrolase; Metalloprotease; Zinc; Complete proteome.
CC FT METAL 112 112 ZINC (POTENTIAL).
CC FT METAL 116 116 ZINC (POTENTIAL).
CC SQ SEQUENCE 387 AA; 42725 MW; AB974E50F138591D CRC64;

Query Match 47.2%; Score 42; DB 1; Length 387;
Best Local Similarity 42.9%; Pred. No. 24;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 KODREDDFDWNPAD 14
Db 279 KSFRKDFENWKELE 292

```

Search completed: May 21, 2004, 12:31:52
 Job time : 2.52817 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 21, 2004, 12:27:57 ; Search time 1.90141 seconds
(without alignments)

2489.088 Million cell updates/sec

Title: US-09-981-649a-24_COPY_412_426

Perfect score: 89

Sequence: 1 KQDREDDFDWNPADR 15

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_virus.*
- 16: sp_bacteriaph.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	89	100.0	474	Q8WYG3	Q8WYG3 homo sapien
2	89	100.0	553	Q8NBV0	Q8NBV0 homo sapien
3	89	100.0	553	Q9NZL7	Q9NZL7 homo sapien
4	89	100.0	553	Q8IUX8	Q8IUX8 homo sapien
5	89	100.0	554	Q9NY67	Q9NY67 homo sapien
6	89	100.0	558	Q9UFG6	Q9UFG6 homo sapien
7	84	94.4	550	Q9JUZ5	Q9JUZ5 mus musculus
8	60	67.4	544	Q8AVH7	Q8AVH7 xenopus lae
9	48	53.9	521	Q42650	Q42650 schizosacch
10	47	52.8	299	Q9NKE0	Q9NKE0 drosophila
11	47	52.8	498	Q97WS5	Q97WS5 sulfolobus
12	47	52.8	511	Q8L3V9	Q8L3V9 zea mays (m
13	47	52.8	618	Q8LK09	Q8LK09 zea mays (m
14	47	52.8	625	Q84FC2	Q84FC2 myxococcus
15	46.5	52.2	876	Q7TSZ6	Q7TSZ6 mus musculus
16	46	51.7	483	Q8G6T6	Q8G6T6 bifidobacte

17	46	51.7	491	16	Q98BF0
18	46	51.7	1833	10	Q7XBD9
19	45	50.6	463	16	Q824S7
20	45	50.6	651	10	Q39720
21	44	49.4	330	12	Q66673
22	44	49.4	339	10	Q66673
23	44	49.4	369	10	Q94IL0
24	44	49.4	435	10	Q9SAF4
25	44	49.4	450	2	Q93JN4
26	44	49.4	715	16	Q8AAG9
27	44	49.4	926	16	Q8Y2M5
28	44	49.4	1094	13	Q804V4
29	44	49.4	1094	13	Q7ZT67
30	44	49.4	1107	13	Q7ZT18
31	43	48.3	4074	4	Q86Z26
32	43	48.3	426	3	Q6L168
33	43	48.3	457	17	Q97X52
34	43	48.3	499	11	Q9D842
35	42.5	47.8	896	5	Q9H1B8
36	42.5	47.8	282	11	Q8BV89
37	42.5	47.8	501	16	Q89IN3
38	42.5	47.8	629	10	Q9MSL9
39	42	47.2	685	11	Q922U7
40	42	47.2	150	10	Q9FEM4
41	42	47.2	150	10	Q9FEM5
42	42	47.2	169	10	Q7X912
43	42	47.2	218	16	Q8UA34
44	42	47.2	240	10	P93045
45	42	47.2	240	10	Q9LUN8
			270	10	Q24347

ALIGNMENTS

RESULT 1

Q8WYG3	Q8WYG3	PRELIMINARY;	PRT;	474 AA.
AC	Q8WYG3;			
DT	01-MAR-2002 (TREMBLrel. 20, Created)			
DT	01-MAR-2002 (TREMBLrel. 20, Last sequence update)			
DE	01-OCT-2003 (TREMBLrel. 25, Last annotation update)			
DE	Hypothetical protein.			
GN	Pp648.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Gu J.R., Wan D.F., Zhao X.T., Zhou X.M., Jiang H.Q., Zhang P.P.,			
RA	Qin W.X., Huang Y., Qiu X.K., Qian L.F., He L.P., Li H.N., Yu Y.,			
RA	Yu J., Han L.H.;			
RT	"Novel human cDNA clones with function of inhibiting cancer cell growth";			
RL	Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.			
CC	-1- SIMILARITY: CONTAINS 1 MAM DOMAIN.			
DR	EMBL; AF193055; AAG22483.1; -			
DR	GO; GO:0016020; C:membrane; IEA.			
DR	GO; GO:0005509; F:calcium ion binding; IEA.			
DR	InterPro; IPR000152; Asx hydroxyl_S.			
DR	InterPro; IPR001881; EGF_Ca			
DR	InterPro; IPR006209; EGF-like.			
DR	InterPro; IPR000998; MAM_domain.			
DR	Pfam; PF00008; EGF; 4.			
DR	Pfam; PF00629; MAM; 1.			
DR	SMART; SM00179; EGF_CA; 3.			
DR	PROSITE; PS00010; ASX_HYDROXYL; 3.			
DR	PROSITE; PS01186; EGF_2; 1.			
DR	PROSITE; PS01187; EGF_CA; 3.			
DR	PROSITE; PS50060; MAM_2; 1.			
KW	Hypothetical protein; EGF-like domain; Glycoprotein.			
SEQUENCE	474 AA; 53144 MW; 5E116F0DF91AC88D CRC64;			

Q98BF0 rhizobium 1
Q7XBD9 zea mays (m
Q824S7 chlamydomoni
Q39720 euglena gra
Q66673 equine herp
Q94IL0 oryza sativ
Q9SAF4 arabidopsis
Q93JN4 rhizobium 1
Q8AAG9 bacteroides
Q8Y2M5 anabaena sp
Q804V4 xenopus lae
Q7ZT67 xenopus lae
Q7ZT18 xenopus lae
Q86Z26 homo sapien
Q6L168 saccharomyc
Q97X52 sulfolobus
Q9D842 mus musculu
Q9H1B8 caenorhabdi
Q8BV89 mus musculu
Q89IN3 bradyrhizob
Q9MSL9 zea mays (m
Q922U7 mus musculu
Q9FEM4 betula verr
Q9FEM5 betula verr
Q7X912 oryza sativ
Q8UA34 agrobacteri
P93045 arabidopsis
Q9LUN8 arabidopsis
Q24347 secale cere

```

Query Match      100.0%; Score 89; DB 4; Length 474;
Best Local Similarity 100.0%; Pred. No. 4.7e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KQDREDDFDWNPADR 15
DB 333 KQDREDDFDWNPADR 347

RESULT 2
Q9NZL7 Q9NZL7 PRELIMINARY; PRT; 553 AA.
ID Q9NZL7;
AC Q9NZL7;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein FLJ97073.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC Tissue=Placenta;
RA Isogai T., Ota T., Nishikawa T., Hayashi K., Otsuki T., Sugiyama T.,
RA Suzuki Y., Nagai K., Sugano S., Ishii S., Kawai-Hio Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Kojima S., Nagahara K.,
RA Masuho Y., Ono T., Okano K., Yoshikawa Y., Aotsuka S., Sasaki N.,
RA Hattori A., Okumura K., Iwayanagi T., Nihomiya K.;
RA "NEDO human cDNA sequencing project.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: CONTAINS 1 MAM DOMAIN.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR InterPro; IPR001881; EGF_CA.
DR InterPro; IPR006210; IEGF.
DR Pfam; PF00008; EGF; 4.
DR Pfam; PF00629; MAM; 1.
DR SMART; SM00179; EGF_CA; 3.
DR PROSITE; PS00010; ASX_HYDROXYL; 3.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 2.
DR PROSITE; PS01187; EGF_CA; 3.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS00060; MAM_2; 1.
DR EGF-like domain; Glycoprotein.
KW SEQUENCE 553 AA; 61318 MW; 3AE93A013CED5880 CRC64;

Query Match      100.0%; Score 89; DB 4; Length 553;
Best Local Similarity 100.0%; Pred. No. 5.6e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KQDREDDFDWNPADR 15
DB 412 KQDREDDFDWNPADR 426

RESULT 3
Q9NZL7 Q9NZL7 PRELIMINARY; PRT; 553 AA.
ID Q9NZL7;
AC Q9NZL7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Epidermal growth factor repeat containing protein.
GN EGF_L6.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20079166; PubMed=10610727;
RA Yeung G., Mulero J.J., Berntsen R.P., Loeb D.B., Drmanac R.,
RA Ford J.E.;
RT EGF_L6: expressed in tumor and fetal tissues.";
RL Genomics 62:304-307(1999).
CC -!- SIMILARITY: CONTAINS 1 MAM DOMAIN.
DR EMBL; AF186084; AAF27812.1; -.
DR HSSP; P00736; IAPQ.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR InterPro; IPR000152; Asx_hydroxyl_S.
DR InterPro; IPR001881; EGF_CA.
DR InterPro; IPR006209; EGF_Like.
DR InterPro; IPR000998; MAM_domain.
DR Pfam; PF00008; EGF; 4.
DR Pfam; PF00629; MAM; 1.
DR SMART; SM00179; EGF_CA; 3.
DR PROSITE; PS00010; ASX_HYDROXYL; 3.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 2.
DR PROSITE; PS01187; EGF_CA; 3.
DR PROSITE; PS00060; MAM_2; 1.
DR EGF-like domain; Glycoprotein.
KW SEQUENCE 553 AA; 61314 MW; 2FF55F167857DE50 CRC64;

Query Match      100.0%; Score 89; DB 4; Length 553;
Best Local Similarity 100.0%; Pred. No. 5.6e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KQDREDDFDWNPADR 15
DB 412 KQDREDDFDWNPADR 426

RESULT 4
Q8IUX8 Q8IUX8 PRELIMINARY; PRT; 553 AA.
ID Q8IUX8;
AC Q8IUX8;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Similar to EGF-like domain, multiple 6.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC Tissue=Brain;
RA Strausberg R.;
RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC038587; AAH38587.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR InterPro; IPR000152; Asx_hydroxyl_S.
DR InterPro; IPR001881; EGF_CA.
DR InterPro; IPR006209; EGF_Like.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR000998; MAM_domain.
DR Pfam; PF00008; EGF; 4.
DR Pfam; PF00629; MAM; 1.
DR SMART; SM00181; EGF; 5.
DR SMART; SM00179; EGF_CA; 3.
DR SMART; SM00137; MAM; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; 3.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 2.
DR PROSITE; PS01187; EGF_CA; 3.

```

```

DR PROSITE; PS0060; MAM 2; 1.
SQ SEQUENCE 553 AA; 61317 MW; 3AE93A0362E861E0 CRC64;

Query Match 100.0%; Score 89; DB 4; Length 553;
Best Local Similarity 100.0%; Pred. No. 5.6e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KQREDDFDWNPADR 15
DB 412 KQREDDFDWNPADR 426

RESULT 5
Q9NY67 PRELIMINARY; PRT; 554 AA.
AC Q9NY67;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN W80.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC Tissue=Brain;
RA Diesterhoef A., Lauber J., Mewes H.W., Gassenhuber J., Wiemann S.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: CONTAINS 1 MAM DOMAIN.
DR EMBL; AL117610; CAB56014.1; -.
DR PIR; T17324; T17324.
DR HSP; P00736; IAPQ.
DR Genew; HGNC:3235; EGFL6.
DR GO; GO:0005615; C:extracellular space; TAS.
DR GO; GO:0005178; F:integrin binding; TAS.
DR GO; GO:0007049; P:cell cycle; TAS.
DR InterPro; IPR00152; Asx hydroxyl_S.
DR InterPro; IPR001881; EGF_CA.
DR InterPro; IPR006209; EGF_Like.
DR InterPro; IPR000998; MAM_domain.
DR Pfam; PF00008; EGF; 4.
DR Pfam; PF00629; MAM; 1.
DR SMART; SM00179; EGF_CA; 3.
DR PROSITE; PS00010; ASX HYDROXYL; 3.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 2.
DR PROSITE; PS01187; EGF_CA; 2.
DR PROSITE; PS00060; MAM_2; 1.
DR Hypothetical protein; EGF-like domain; Glycoprotein.
KW NON TER
FT
SQ SEQUENCE 558 AA; 61828 MW; AA38D7DCE402BFA3 CRC64;

Query Match 100.0%; Score 89; DB 4; Length 558;
Best Local Similarity 100.0%; Pred. No. 5.7e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KQREDDFDWNPADR 15
DB 417 KQREDDFDWNPADR 431

RESULT 7
Q9JZ5 PRELIMINARY; PRT; 550 AA.
AC Q9JZ5;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN EGFL6 OR W80.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC Franco B.;
RA Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20241927; PubMed=10777661;
RA Buchner G., Orfanelli U., Quaderi N., Bassi M.T., Andolfi G.;
RT "Identification of a new EGF-repeat-containing gene from human Xp22:
RT Acandidate for developmental disorders.";
RL Genomics 65:16-23(2000).
CC -!- SIMILARITY: CONTAINS 1 MAM DOMAIN.
DR EMBL; AJ245672; CAB92138.1; -.
DR HSP; P35555; 1EMN.
DR MGD; MGI:1858599; Egfl6.

```



```

DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR InterPro; IPR000152; Asx_hydroxyl_S.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR006209; EGF like.
DR InterPro; IPR000998; MAM_domain.
DR Pfam; PF00008; EGF; 4.
DR Pfam; PF00629; MAM; 1.
DR SMART; SM00179; EGF CA; 3.
DR PROSITE; PS00010; ASX_HYDROXYL; 3.
DR PROSITE; PS00022; EGF 1; 1.
DR PROSITE; PS01186; EGF 2; 2.
DR PROSITE; PS01187; EGF CA; 3.
DR PROSITE; PS00060; MAM 2; 1.
DR Hypothetical protein_EGF-like domain; Glycoprotein.
KW SEQUENCE 550 AA; 61520 MW; DE936325C9F31B3 CRC64;

Query Match 94.4%; Score 84; DB 11; Length 550;
Best Local Similarity 93.3%; Pred. No. 3.6e-05;
Matches 14; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KQREDDFDWNPADR 15
DB 409 KQREDDFDWNPADR 423

RESULT 8
Q8AVH7 PRELIMINARY; PRT; 544 AA.
AC Q8AVH7;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Similar to EGF-like domain, multiple 6.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RA Klein S.; Strausberg R.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC042275; AAH42275.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR InterPro; IPR000152; Asx_hydroxyl_S.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR006209; EGF like.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR000998; MAM_domain.
DR Pfam; PF00008; EGF; 3.
DR Pfam; PF00629; MAM; 1.
DR PRINTS; PR00020; MAMDOMAIN.
DR SMART; SM00181; EGF; 5.
DR SMART; SM00179; EGF CA; 3.
DR SMART; SM00137; MAM; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; 3.
DR PROSITE; PS00022; EGF 1; 1.
DR PROSITE; PS01186; EGF 2; 3.
DR PROSITE; PS01187; EGF CA; 3.
DR PROSITE; PS00060; MAM 2; 1.
DR SEQUENCE 544 AA; 61048 MW; 24865A62D2EAB7B5 CRC64;

Query Match 67.4%; Score 60; DB 13; Length 544;
Best Local Similarity 66.7%; Pred. No. 0.26;
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 KQREDDFDWNPADR 15
DB 411 KQSKDDFDWKHAER 425

```

RESULT 9

```

O42650 PRELIMINARY; PRT; 521 AA.
AC O42650;
DT 01-JAN-1999 (TrEMBLrel. 09, Created)
DT 01-JAN-1999 (TrEMBLrel. 09, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN SPAC10F6.10.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972h-;
RA Connor R.; Churcher C.M.; Barrell B.G.; Rajandream M.A.; Wood V.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL009197; CAAL5723.1; -.
DR PIR; T37504; T37504.
DR GeneDB SPombe; SPAC10F6.10; -.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004713; F:protein-tyrosine kinase activity; IEA.
DR GO; GO:0006468; F:protein amino acid phosphorylation; IEA.
DR InterPro; IPR000687; RIO1_UNK.
DR InterPro; IPR008266; Tyr_kinase_AS.
DR Pfam; PF01163; RIO1; 1.
DR SMART; SM00090; RIO; 1.
DR PROSITE; PS00109; PROTEIN KINASE_TYR; 1.
DR PROSITE; PS01245; RIO1; 1.
KW Hypothetical protein.
KW SEQUENCE 521 AA; 60231 MW; CADF05173CB54F63 CRC64;

Query Match 53.9%; Score 48; DB 3; Length 521;
Best Local Similarity 50.0%; Pred. No. 22;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 REDDDFWNPA 13
DB 31 REDDIDWQA 40

RESULT 10
Q9NKB0 PRELIMINARY; PRT; 299 AA.
AC Q9NKB0;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN BG:DS06874.6 OR CG18420.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=99403001; PubMed=10471707;
RA Ashburner M.; Misra S.; Roote J.; Lewis S.E.; Blazej R.; Davis T.;
RA Doyle C.; Galle R.; George R.; Harris N.; Hartzell G.; Harvey D.;
RA Hong L.; Houston K.; Hoskins R.; Johnson G.; Martin C.; Moshrefi A.;
RA Palazzolo M.; Reese M.G.; Spradling A.; Tsang G.; Wan X.; Whitelaw K.;
RA Celniker S.; Rubin G.M.;
RT Drosophila melanogaster: the Adh region.";
RL Genetics 153:179-219 (1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;

```

RA Celniker S.E., Agbayani A., Arcaina T.T., Baxter E., Blazej R.G.,
 RA Butenoff C., Champe M., Chavez C., Chew M., Ciesiolka L., Doyle C.M.,
 RA Farfan D.E., Galle R., George R.A., Harris N.L., Hoskins R.A.,
 RA Houston K.A., Hummasti S.R., Karra K., Kearney L., Kim E., Lee B.,
 RA Lewis S., Li P., Lomotan M.A., Mazda P., Moshrefi A.R., Moshrefi M.,
 RA Nixon K., Pacleb J.M., Park S., Pfeiffer B., Poon L., Sequeira A.,
 RA Sethi H., Snir E., Svirskas R.R., Wan K.H., Weinburg T., Zhang R.,
 RA Zieran L.L., Rubin G.M.,
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
 CC EMBL: AE003411; AAF44895.1; -;
 DR HSPSP; P00734; 1B7X.
 DR FlyBase; FBgn028866; BG:DS06874.6.
 DR GO: GO:004263; F:chymotrypsin activity; IEA.
 DR GO: GO:0008233; F:peptidase activity; IEA.
 DR GO: GO:0004295; F:trypsin activity; IEA.
 DR GO: GO:0006508; F:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR009003; Cys_Ser_trypsin.
 DR InterPro; IPR001254; Peptidase S1.
 DR InterPro; IPR001314; Peptidase_S1A.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00020; TRY_PSC; 1.
 DR PROSITE; PS0240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; 1.
 DR KW Hypothetical protein; Hydrolase; Protease; Serine protease.
 DR SEQUENCE 299 AA; 33696 MW; 1FC372E5D5ACD987 CRC64;
 Query Match 52.8%; Score 47; DB 5; Length 299;
 Best Local Similarity 63.6%; Pred. No. 17;
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 QY 1 KODREDDFDWN 11
 Db 282 KPDKEPFDWN 292
 RESULT 11
 Q97WS5 PRELIMINARY; PRT; 498 AA.
 AC Q97WS5;
 DT 01-OCT-2001 (TREMBlrel. 18, Created)
 DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Acetyl-CoA synthetase (Acetate-CoA ligase) (acaA-5) (BC 6.2.1.1).
 GN ACSA-5 OR SS02041.
 OS Sulfolobus solfataricus.
 OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
 CC Sulfolobus.
 CC NCBI_TaxID=2287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 35092 / DSM 1617 / P2;
 RX MEDLINE=2332296; PubMed=11427726;
 RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
 RA Awayez M.J., Chan-Weher C.C.-Y., Clausen I.G., Curtis B.A.,
 RA De Moors A., Erauso G., Fletcher C., Gordon P.M.K.,
 RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
 RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
 RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
 RA Garrett R.A., Ragan M.A., Sengen C.W., Van der Oost J.;
 RA "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
 RT Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
 RL EMBL: AE006811; AAK42227.1; -;
 DR FTR; D90371; D90371.
 DR GO: GO:0016020; C:membrane; IEA.
 DR GO: GO:0003987; F:acetate-CoA ligase activity; IEA.
 DR GO: GO:0016874; F:ligase activity; IEA.
 DR GO: GO:0008233; F:peptidase activity; IEA.
 DR GO: GO:0008152; P:metabolism; IEA.
 DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000873; AMP-bind.
 DR InterPro; IPR000508; Peptidase_S26.

DR Pfam; PF00501; AMP-binding; 2.
 DR PROSITE; PS00455; AMP_BINDING; 1.
 DR PROSITE; PS00761; SPASE_I_3; 1.
 KW Ligase; Complete proteome.
 SQ SEQUENCE 498 AA; 57091 MW; A6D57E5C1FAE3B4B CRC64;
 Query Match 52.8%; Score 47; DB 17; Length 498;
 Best Local Similarity 72.7%; Pred. No. 30;
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 4 REDDFDWNPAD 14
 Db 199 RENDVFNWPNAD 209
 RESULT 12
 Q8L3V9 PRELIMINARY; PRT; 511 AA.
 ID Q8L3V9;
 AC Q8L3V9;
 DT 01-OCT-2002 (TREMBlrel. 22, Created)
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Histone deacetylase HDAl10 isoform 2 (Histone deacetylase HDAl10 isoform 3).
 DE HDAl10.
 GN HDAl10.
 OS Zea mays (Maize).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 CC PACCAD clade; Panicoideae; Andropogoneae; Zea.
 CC NCBI_TaxID=4577;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. B73;
 RA Chandler V.L., Cone K.C., Kaeppler H.F., Kaeppler S.M.;
 RT "Sequences from the Plant Chromatin Consortium (NSF Plant Genome Project 9975930).";
 RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. B73;
 RA Springer N.M., Schmitt L.T., Bergstrom D., Selinger D.A., Guthrie E.,
 RA Kaeppler S.M.;
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF527612; AAM93213.1; -;
 DR EMBL; AF527613; AAM93214.1; -;
 DR InterPro; IPR000286; His_deacetylase.
 DR InterPro; IPR001876; ZnF_RanGDP.
 DR Pfam; PF00850; Hist_deacetyl; 1.
 DR PRINTS; PR01270; HDASUPER.
 DR PROSITE; PS01358; ZF_RANBP2_1; 1.
 DR SEQUENCE 511 AA; 55031 MW; E45B0302A0D7E217 CRC64;
 Query Match 52.8%; Score 47; DB 10; Length 511;
 Best Local Similarity 53.8%; Pred. No. 31;
 Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 QY 3 REDDFDWNPADR 15
 Db 25 BEDDSDWEPASR 37
 RESULT 13
 Q8LK09 PRELIMINARY; PRT; 618 AA.
 ID Q8LK09;
 AC Q8LK09;
 DT 01-OCT-2002 (TREMBlrel. 22, Created)
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Histone deacetylase HDAl10 isoform 1.
 GN HDAl10.
 OS Zea mays (Maize).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

OC PACCAD clade; Panicoideae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. B73;
RA Chandler V.L., Cone K.C., Kaeppler H.F., Kaeppler S.M.;
RT "Sequences from the Plant Chromatin Consortium (NSF Plant Genome
Project 9975930).";
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. B73;
RA Springer N.M., Schmitt L.T., Bergstrom D., Selinger D.A., Guthrie E.,
RA Kaeppler H.F., Kaeppler S.M.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF527611; AAM91212.1; -.
DR InterPro; IPR00286; His deacetylase.
DR InterPro; IPR001876; ZnF RANGDP.
DR Pfam; PF00850; Hist deacetyl; 1.
DR PRINTS; PR01270; HDASUPER.
DR PROSITE; PS01358; ZF_RANBP2_1; 1.
SQ SEQUENCE 618 AA; 66058 MW; 8395CFA75F79DE63 CRC64;
Query Match 52.8%; Score 47; DB 10; Length 618;
Best Local Similarity 53.8%; Pred. No. 38;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
QY 3 DREDDFDWNPADR 15
Db 132 EDDDSWEPASR 144
RESULT 14
ID Q84FC2 PRELIMINARY; PRT; 625 AA.
AC Q84FC2;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative histidine kinase.
OS Myxococcus xanthus.
OC Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
OC Cytophactereae; Myxococcaceae; Myxococcus.
OX NCBI_TaxID=34;
RN [1]
RP SEQUENCE FROM N.A.
RA Hartzell P.L., Youderian P.A.;
RT "Identification of genes required for adventurous gliding motility in
RT Myxococcus xanthus with the transposable element mariner.";
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY204469; AAC22893.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0016301; F:kinase activity; IEA.
DR GO; GO:0000155; F:two-component sensor molecule activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR GO; GO:0007165; P:signal transduction; IEA.
DR InterPro; IPR003018; GAF.
DR InterPro; IPR003661; His_kinA_N.
DR InterPro; IPR001610; PAC.
DR InterPro; IPR000014; PAS_domain.
DR Pfam; PF01590; GAF; 1.
DR Pfam; PF00512; HiskA; 1.
DR Pfam; PF00785; PAC; 2.
DR Pfam; PF00989; PAS; 1.
DR SMART; SM00065; GAF; 1.
DR SMART; SM00388; HiskA; 1.
DR TIGRfams; TIGR00229; sensory_box; 2.
DR PROSITE; PS50112; PAS; 1.
KW Kinase.
SQ SEQUENCE 625 AA; 68154 MW; C3070D65E0DD5958 CRC64;
Query Match 52.8%; Score 47; DB 2; Length 625;

Best Local Similarity 63.6%; Pred. No. 39;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 4 REDDFDWNPAD 14
Db 571 KEDDFAWKPYD 581
RESULT 15
ID Q7TSZ6 PRELIMINARY; PRT; 876 AA.
AC Q7TSZ6;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Kpn1 protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=22389257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heish F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettner M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalls D.E., Scherch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Strausberg R.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC052711; AAH52711.1; -.
SQ SEQUENCE 876 AA; 97167 MW; 6A11DCE03086821 CRC64;
Query Match 52.2%; Score 46.5; DB 11; Length 876;
Best Local Similarity 69.2%; Pred. No. 68;
Matches 9; Conservative 1; Mismatches 2; Indels 1; Gaps 1;
QY 1 KQD-REDDFDWNP 12
Db 332 KQDENDDDDDWNP 344

Search completed: May 21, 2004, 12:33:18
Job time : 4.90141 secs